

BỘ GIÁO DỤC VÀ ĐÀO TẠO

VIỆN HÀN LÂM KHOA HỌC
VÀ CÔNG NGHỆ VIỆT NAM

HỌC VIỆN KHOA HỌC VÀ CÔNG NGHỆ

Trần Thị Huyền Trang

NGHIÊN CỨU ĐA HÌNH MỘT SỐ GEN LIÊN QUAN ĐẾN
TÍNH TRẠNG TĂNG TRƯỞNG Ở CÁ TRA NUÔI
(*Pangasianodon hypophthalmus*)

Chuyên ngành: Công nghệ sinh học
Mã số: 9 42 02 01

TÓM TẮT LUẬN ÁN TIẾN SĨ SINH HỌC

Hà Nội – Năm 2023

Công trình được hoàn thành tại: Học viện Khoa học và Công nghệ -
Viện Hàn lâm Khoa học và Công nghệ Việt Nam.

Người hướng dẫn khoa học: TS. Kim Thị Phương Oanh

Phản biện 1: ...

Phản biện 2: ...

Phản biện 3:

Luận án sẽ được bảo vệ trước Hội đồng đánh giá luận án tiến sĩ cấp Học viện, họp tại Học viện Khoa học và Công nghệ - Viện Hàn lâm Khoa học và Công nghệ Việt Nam vào hồi ... giờ ..', ngày ... tháng ... năm 2023

Có thể tìm hiểu luận án tại:

- Thư viện Học viện Khoa học và Công nghệ

- Thư viện Quốc gia Việt Nam

MỞ ĐẦU

1. Tính cấp thiết của luận án

Cá tra nuôi (*Pangasianodon hypophthalmus*) là loài cá của vùng lưu vực sông Mê Kông và sông ChaoPhraya (Thái Lan), được nuôi phổ biến và có giá trị kinh tế rất lớn ở Việt Nam và một số nước khác. Theo quyết định 50/2018/QĐ-TTg của Thủ tướng chính phủ, cá tra là đối tượng nuôi chủ lực ở nước ta phục vụ xuất khẩu và tiêu thụ trong nước. Tuy nhiên nghề nuôi cá tra của Việt Nam đang gặp nhiều thách thức lớn như chất lượng giống ngày càng giảm sút, hàng năm dịch bệnh xảy ra làm thiệt hại lớn cho người nuôi. Để phát triển bền vững ngành sản xuất cá tra, một trong những nhiệm vụ cấp thiết cần thực hiện đầu tiên chính là nâng cao chất lượng nguồn giống. Trong đó, hướng tìm kiếm các chỉ thị phân tử, đặc biệt là các SNP trên gen đích liên quan đến các tính trạng tăng trưởng rất tiềm năng. Hệ thống các yếu tố tăng trưởng tương tự insulin (Insulin-like Growth Factor (IGF)) bao gồm các phối tử IGF1, IGF2, các thụ thể tương ứng của IGF (IGF Receptor (IGFR)) và các protein bám vào các IGF (IGF binding proteins (IGFBPs)). Sự tương tác giữa IGF, IGFR và IGFBP cùng các protein khác trong chuỗi dẫn truyền tín hiệu sẽ dẫn tới quá trình sinh trưởng, biệt hóa và tăng sinh tế bào ở động vật có xương sống nói chung và cá xương nói riêng. Do đó, nghiên cứu đa hình một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng ở cá tra nuôi (*P. hypophthalmus*) là rất cần thiết, góp phần tìm ra chỉ thị phân tử hỗ trợ chọn giống cá tra nuôi theo hướng tăng trưởng nhanh, phục vụ nhu cầu của ngành nuôi trồng thủy sản ở nước ta.

2. Mục tiêu nghiên cứu của luận án

Nghiên cứu phát hiện các đa hình SNP trên một số gen thuộc hệ thống IGF và kiểm nghiệm sự liên quan của các SNP này với tính trạng tăng trưởng ở cá tra nuôi (*Pangasianodon hypophthalmus*), đề xuất được một số chỉ thị phân tử tiềm năng hướng tới ứng dụng trong chương trình chọn giống cá tra theo hướng tăng trưởng.

3. Các nội dung nghiên cứu chính của luận án

- Phân tích cấu trúc các gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 thuộc hệ thống IGF của cá tra nuôi, tìm các vùng quan trọng để xác định đa hình SNP và giải trình tự các vùng đó trên 1 cá thể bằng phương pháp Sanger để kiểm tra tính xác thực về trình tự so với trình tự tham chiếu.

- Giải trình tự các vùng quan trọng thuộc 9 gen nêu trên trên bộ mẫu khởi tạo gồm 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm bằng phương pháp Sanger, so sánh các trình tự tương ứng thu được với trình tự tham chiếu để phát hiện SNP trên các gen và sàng lọc các SNP.

- Xác định kiểu gen (genotyping) các SNP được sàng lọc trên bộ mẫu kiểm nghiệm gồm 70 cá thể tăng trưởng nhanh và 70 cá thể tăng trưởng chậm bằng phương pháp kéo dài một nucleotide (SBE).

- Phân tích dữ liệu SNP thu được từ cả bộ mẫu khởi tạo và bộ mẫu kiểm nghiệm, gồm tổng cộng 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm, phân tích sự liên quan của các SNP được sàng lọc, các haplotype, tổ hợp kiểu gen liên quan với tính trạng tăng trưởng của cá tra nuôi.

CHƯƠNG 1. TỔNG QUAN

1.1. Giới thiệu chung về cá tra nuôi *Pangasianodon hypophthalmus*

Loài cá tra nuôi *Pangasianodon hypophthalmus* (Sauvage, 1878) có tên gốc là *Helicophagus hypophthalmus* (Sauvage, 1878), hay còn có tên khác là *Pangasius sutchi*, thuộc chi *Pangasius*, họ cá tra (*Pangasiidae*), bộ cá da trơn hay cá nheo (*Siluriformes*), là một trong những loài cá được nuôi phổ biến của vùng lưu vực sông Mê Kông (Việt Nam, Thái Lan, Lào, Campuchia) và sông ChaoPhraya (Thái Lan). Căn cứ Quyết định số 50/2018/QĐ-TTg của Thủ tướng chính phủ, cá tra là đối tượng thủy sản nuôi chủ lực ở nước ta, phục vụ xuất khẩu và tiêu thụ trong nước. Trong năm 2022, ngành cá tra dự kiến kim ngạch xuất khẩu đạt trên 1,6 tỷ USD. Để đạt được mục tiêu đề ra, Bộ Nông nghiệp và Phát triển Nông thôn yêu cầu các địa phương nâng cao chất lượng giống cá tra, chỉ đạo sản xuất cung ứng đủ con giống chất lượng cao để nâng cao hiệu quả sản xuất, giảm hao hụt, hạ giá thành sản xuất. Năm 2018, toàn bộ hệ gen của cá tra nuôi (*P. hypophthalmus*) đã được giải mã bằng NGS với kích thước hệ gen nhân khoảng 700 Mbp, chứa 28600 gen mã hóa protein, có thể được sử dụng làm hệ gen tham chiếu, trên cơ sở đó phát triển các nghiên cứu tìm kiếm chỉ thị phân tử liên quan đến các tính trạng quan tâm phục vụ chọn giống.

1.2. Tăng trưởng (phát triển cơ) ở cá và các gen liên quan

Phát triển cơ là một trong những quá trình tăng trưởng quan trọng ở cá được điều khiển bởi hệ trục hormone liên quan đến sinh trưởng, trong đó có hệ thống các nhân tố tăng trưởng tương tự insulin (IGF). Nhiều

ngiên cứu đã chỉ ra mối liên quan giữa đa hình ở các gen thuộc hệ trục hormone liên quan đến tăng trưởng, đặc biệt là các gen thuộc hệ thống IGF và tính trạng tăng trưởng của cá xương.

1.3. Vai trò, chức năng và đặc điểm cấu trúc của một số gen /protein thuộc hệ thống IGF

Hệ thống các yếu tố tăng trưởng tương tự insulin (Insulin-like Growth Factor (IGF)) bao gồm các phối tử IGF1, IGF2, các thụ thể tương ứng của IGF (IGF Receptor (IGFR)) và các protein bám vào các IGF (IGF binding proteins (IGFBPs)). Sự tương tác giữa IGF, IGFR và IGFBP cùng các protein khác trong chuỗi dẫn truyền tín hiệu sẽ dẫn tới quá trình sinh trưởng, biệt hóa và tăng sinh tế bào. IGF1 và IGF2 thúc đẩy sự tăng trưởng bằng cách tăng cường sự phân chia và biệt hóa các tế bào vệ tinh cơ xương, kích thích sinh tổng hợp protein và phát triển phình cơ, đồng thời ức chế phân hủy protein và teo cơ. Phối tử IGF truyền tín hiệu thông qua IGF1R – một thụ thể tyrosine kinase. IGF2R có khả năng bám cao hơn với IGF2 nhưng theo hướng để phân hủy IGF2. Các IGFBP điều khiển hoạt tính sinh học của các IGF trong môi trường ngoại bào, từ đó tác động đến khả năng bám của các phối tử IGF với các thụ thể. Các IGFBPs thuộc họ protein tiết, được chia làm 6 loại từ IGFBP-1 đến -6, có ái lực cao với IGF. Ngoài ra, IGFBP-7 hay còn gọi là IGFBP-rP1 (thuộc 9 loại protein liên quan với IGFBP (IGFBP related proteins (IGFBP-rPs)) cũng có cấu trúc tương đồng với IGFBP, nhưng có ái lực với IGF thấp hơn các IGFBP-1 đến -6.

Trong khi cấu trúc gen *IGF1* ở lớp thú có 6 exon thì ở lớp cá, gen này chỉ có 5 exon. Cấu trúc gen *IGF2* ở các loài cá khi được so sánh với cấu trúc gen *IGF2* của lớp thú cho thấy sự ít phức tạp hơn, ổn định với 4 exon. Cấu trúc gen *IGF1Ra* và *IGF1Rb* ở các loài cá gồm 21 exon, cách nhau bởi các intron với kích thước đa dạng. Các gen *IGFBP* từ 1 đến 6 đều chứa 4 exon quy định các IGFBP có chung cấu trúc bảo thủ với hai domain chức năng là IGFBP ở đầu N và thyroglobulin tuýp 1 ở đầu C, được nối với nhau bằng domain nối ở giữa.

1.4. Cơ sở tiến hành xác định các đa hình đơn nucleotit (SNP) trên một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng của cá tra nuôi

Vai trò quan trọng của SNP trong việc xác định các chỉ thị phân tử liên quan đến tính trạng quan tâm cả ở quy mô hệ gen lẫn quy mô gen đích

đều đã được chứng minh. Trong thủy sản, SNP thường được sử dụng rộng rãi trong việc xác định các chỉ thị phân tử tiềm năng liên quan đến các tính trạng chọn giống, kinh tế như tăng trưởng, chống chịu bệnh. Ở quy mô tìm kiếm SNP tiềm năng trên gen đích, các phương pháp SNP genotyping dựa trên nguyên lý PCR sử dụng đầu dò, hay giải trình tự đều khá phát triển, có thể được ứng dụng rộng rãi. Trong đó, phương pháp kéo dài một nucleotit (Single Base Extension –SBE) có thể phát hiện tetra – allelic SNPs với độ nhạy và đặc hiệu cao và khả năng giải trình tự tự động trong thời gian ngắn. Các nghiên cứu về SNP trên một số gen thuộc hệ thống IGF ở cá xương cũng đã chỉ ra mối liên quan giữa các SNP trên các gen *IGF1*, *IGF2*, *IGF1R* cũng như các kiểu gen đơn bội (haplotype), tổ hợp kiểu gen phát sinh với tính trạng tăng trưởng của các loài cá như: cá hồi Đại Tây Dương *Salmo salar*, cá chép *Cyprinus carpio*, cá vược châu Âu *Dicentrarchus labrax*, cá rô phi sông Nin *Oreochromis niloticus*, cá bản địa Trung Quốc *Odontobutis potamophila*. Từ hệ gen tham chiếu, các gen thuộc hệ thống IGF của cá tra nuôi đã được xác định gồm 2 gen quy định cho IGF1, 2 gen quy định cho IGF2, 4 gen quy định IGF1R, 11 gen quy định IGFBP gồm IGFBP-1, -2, -3, -5, -6, -7, có thể được sử dụng để thiết kế các thí nghiệm phân tích đa hình trên gen liên quan đến tính trạng quan tâm.

CHƯƠNG 2. NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP

2.1. Nguyên vật liệu

Các cá thể cá tra nuôi (*P. hypophthalmus*) sử dụng trong luận án được kế thừa từ các đề tài nghiên cứu trước đây. Cụ thể, 160 cá thể đã được chọn lọc từ thế hệ thứ 3 (G3) cộng gộp (năm 2015) của quần đàn chọn giống cá tra theo hướng tăng trưởng dựa trên phương pháp di truyền số lượng của Viện Nghiên cứu Nuôi trồng Thủy sản 2, Việt Nam. Trong đó, 20 cá thể thuộc bộ mẫu khởi tạo được sử dụng để phát hiện và sàng lọc SNP gồm: 10 cá thể tăng trưởng nhanh có giá trị chọn giống ước đoán (EBV) cá thể cao nhất thuộc 9 gia đình có EBV gia đình cao nhất và 10 cá thể tăng trưởng chậm có EBV cá thể thấp nhất thuộc 9 gia đình có EBV gia đình thấp nhất. Bộ mẫu kiểm nghiệm gồm 140 cá thể sử dụng để xác định các SNP được sàng lọc trên quy mô lớn hơn gồm: 70 cá thể tăng trưởng nhanh có EBV cá thể cao nhất thuộc 24 gia đình có EBV gia đình cao nhất theo thứ tự xếp hạng và 70 cá thể tăng trưởng chậm có EBV cá thể thấp nhất thuộc 31 gia đình có EBV gia đình thấp nhất theo thứ tự xếp hạng.

2.2. Phương pháp

2.2.1. Phân tích cấu trúc của các gen đích

Dựa trên trình tự tham chiếu của các gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 của cá tra nuôi đã được giải mã bằng NGS, cấu trúc các gen/protein suy diễn được phân tích bằng phần mềm BLASTn, ProteinBLAST (blastp), SignalP5.0 để xác định các vùng gen quan trọng quy định domain chức năng, chuỗi peptit tín hiệu trên protein, các vùng điều hòa phiên mã, dịch mã để giải mã trên nhiều cá thể nhằm phát hiện SNP.

2.2.2. Tách chiết DNA tổng số

DNA tổng số từ các tế bào mô vây được tách chiết theo phương pháp thường quy sử dụng phenol: chloroform, sau đó được kiểm tra nồng độ và chất lượng bằng quang phổ kế NanoDrop One Spectrophotometer (Thermo Fisher Scientific) và điện di trên gel agarose 1 %.

2.2.3. Khuếch đại các gen bằng phản ứng PCR

Để khuếch đại các đoạn gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 đã xác định, DNA tổng số tách chiết từ các mẫu vây cá tra được dùng làm khuôn cho phản ứng PCR với 55 cặp mồi thiết kế bằng phần mềm Primer 3 (v.0.4.0). Sản phẩm PCR sau đó được điện di kiểm tra kích thước trên gel agarose 1% và tinh sạch bằng kit Thermo Scientific GeneJET PCR Purification Kit (Thermo Fisher Scientific) theo hướng dẫn của nhà sản xuất.

2.2.4. Giải trình tự các gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 bằng phương pháp Sanger

Các sản phẩm PCR tinh sạch được sử dụng làm khuôn cho phản ứng PCR giải trình tự sử dụng bộ kit BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific) theo hướng dẫn của nhà sản xuất. Các sản phẩm PCR giải trình tự được tinh sạch và điện di mao quản trên hệ thống ABI@3500 Genetic Analyzer (Applied Biosystems). Kết quả giải mã thu được từ máy giải trình tự được biên tập bằng phần mềm BioEdit.

2.2.5. Kiểm tra tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham chiếu

Trình tự của các gen đích thuộc hệ thống IGF của một cá thể cá tra nuôi giải mã bằng Sanger được so sánh với các trình tự tham chiếu bằng

phần mềm BLASTn để tìm ra mức độ tương đồng, phản ánh tính xác thực của trình tự giải mã bằng Sanger so với trình tự tham chiếu.

2.2.6. Phát hiện và sàng lọc các SNP trên các gen đích trên bộ mẫu khởi tạo

Để phát hiện SNP, trình tự nucleotit của các đoạn được xác định ở 20 cá thể thuộc bộ mẫu khởi tạo được giải mã bằng phương pháp Sanger, sau đó so sánh với trình tự tham chiếu tương ứng giải mã bằng NGS thông qua phần mềm MUSCLE.

Trong số các SNP được phát hiện, những SNP được đánh giá và sàng lọc sẽ là SNP được xác định ở cả 20 cá thể (số lượng cá thể không xác định được SNP (NN) = 0), gây nên sự thay đổi axit amin trên protein tương ứng ⁽¹⁾, hoặc đảm bảo có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên ⁽²⁾ và thành phần kiểu gen và/hoặc thành phần alen giữa 2 nhóm khác biệt có ý nghĩa thống kê ^{(3), (4)} (giá trị $p < 0,05$ theo kiểm định Fisher Exact Test). Tần số xuất hiện SNP của mỗi nhóm được xác định bằng tỉ lệ giữa số cá thể mang kiểu gen chứa alen thiểu số và số cá thể mang kiểu gen chứa alen còn lại. Tỉ lệ này cần đạt ít nhất 0,3 để đảm bảo SNP đó không phải chỉ xuất hiện ở số ít cá thể riêng lẻ mà phổ biến trong quần thể. Alen thiểu số được xác định là alen có tần số nhỏ hơn trong tổng số alen của 20 cá thể. Sự khác biệt về thành phần kiểu gen/ thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm được xác định bằng phần mềm SHEsis.

2.2.7. Xác định các SNP được sàng lọc trên bộ mẫu kiểm nghiệm bằng phương pháp kéo dài một nucleotit (SBE)

Các SNP được sàng lọc được chia làm 2 nhóm để xác định bằng phản ứng SBE multiplex trên 140 cá thể thuộc bộ mẫu kiểm nghiệm nhằm giảm thiểu số lượng phản ứng. Ngoài ra, bộ kit ABI SNaPshot Multiplex PCR Kit (Applied BioSystems) được sử dụng để tiến hành phản ứng SBE multiplex cho phép xác định tối đa 10 SNP trong một phản ứng. Các môi SBE được thiết kế chia làm 2 nhóm, mỗi môi trong nhóm đều chứa phần trình tự chính bắt cặp bổ sung với trình tự liền trước đầu 3' của SNP và phần trình tự lặp ở đuôi có kích thước khác nhau giữa các môi trong nhóm. DNA tổng số tách chiết từ mẫu vây của 70 cá thể tăng trưởng nhanh và 70 cá thể tăng trưởng chậm được sử dụng làm khuôn cho phản ứng PCR nhân các đoạn gen chứa các SNP đã được sàng lọc. Các cặp môi cho phản ứng PCR được thiết kế lại để có thể thu được các đoạn chứa SNP có kích thước

đồng đều. Các đoạn sản phẩm PCR chứa các SNP thuộc cùng một nhóm của mỗi cá thể được gộp lại, tinh sạch bằng kit GeneJET PCR Purification Kit (Thermo Fisher Scientific) theo hướng dẫn của nhà sản xuất. Phản ứng SBE được tiến hành bằng bộ kit ABI SNaPshot Multiplex PCR Kit (Applied BioSystems) theo hướng dẫn của nhà sản xuất, sản phẩm thu được được tinh sạch bằng enzyme Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific) và được giải trình tự bằng hệ thống ABI3500 Applied BioSystems. Kết quả SNP genotyping được xử lý bằng phần mềm GeneMapper 5.1 (Applied BioSystems).

2.2.8. Phân tích dữ liệu

Dữ liệu SNP genotyping thu được từ 160 cá thể của bộ mẫu khởi tạo và bộ mẫu kiểm nghiệm gồm 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm được sử dụng để: đánh giá đa dạng di truyền thông qua thông tin đa hình (Polymorphism Information Content (PIC)) và tần số alen thiểu số (Minor Allele Frequency (MAF)) bằng phần mềm GeneCalc; đánh giá khả năng di truyền cùng nhau của các SNP được sàng lọc trên cùng một gen thông qua sự mất cân bằng liên kết (linkage disequilibrium – LD) bằng phần mềm SHEsis; tìm kiếm các chỉ thị SNP tiềm năng qua đánh giá sự khác biệt có ý nghĩa về thành phần kiểu gen, thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm; đánh giá tác động tổng hợp của các SNP lên tính trạng tăng trưởng bằng: sự khác biệt có ý nghĩa giữa hai nhóm về tần số kiểu gen đơn bội haplotype tạo nên từ các SNP trên cùng một gen hay từ các chỉ thị SNP tiềm năng, sự khác biệt có ý nghĩa giữa hai nhóm về tần số tổ hợp kiểu gen tạo nên từ các SNP tiềm năng, các sự khác biệt có ý nghĩa được kiểm định Fisher Exact Test với giá trị $p < 0,05$ bằng phần mềm SHEsis và phần mềm NCSS 2021 Data Analysis (NCSS Statistic Software).

CHƯƠNG 3. KẾT QUẢ

3.1. Phân tích cấu trúc của một số gen/protein thuộc hệ thống IGF của cá tra nuôi

Tất cả các gen thuộc hệ thống IGF của cá tra nuôi được phân tích trong luận án đều có mức độ tương đồng về trình tự đạt cao nhất với các trình tự tương ứng của cá nheo Mỹ *Ictalurus punctatus*.

Trình tự gen *IGF1* dài khoảng 19,4 kb chứa 4 exon đều tham gia vào mã hóa cho protein IGF1. Protein IGF1 có chuỗi peptit tín hiệu dài khoảng 33 axit amin, cấu trúc domain IGF chứa 6 phân tử cysteine nằm rải rác và

các motif bám với IGFBP, thụ thể tuýp 1 và tuýp 2 của IGF/Insulin. Các vùng được khuếch đại để tìm kiếm các SNP trên gen *IGF1* gồm: exon 1, 2 tham gia mã hóa cho chuỗi peptit tín hiệu và domain IGF1 và các vùng trình tự intron xung quanh, một phần trình tự exon 4 có chứa codon kết thúc và vùng 3'-UTR cùng intron liền trước. Trình tự gen *IGF2* dài khoảng 4,9 kb chứa 4 exon đều tham gia vào mã hóa cho các phần thuộc protein IGF2 với cấu trúc domain IGF tương tự IGF1. Do có kích thước nhỏ, nên toàn bộ phần trình tự thuộc 4 exon tham gia mã hóa cho protein và các intron liền kề của gen *IGF2* đều được khuếch đại để tìm kiếm SNP. Gen *IGF1R* dài 93,9 kb, có 21 exon, với vùng mã hóa protein bắt đầu từ exon 1 và kết thúc ở exon 21. Protein IGF1R gồm có các domain chức năng theo thứ tự từ đầu N đến đầu C là: domain thụ thể L thứ nhất, domain tương tự Furin giàu cysteine, domain thụ thể L thứ 2, ba domain Fibronectin tuýp 3 và domain Tyrosine kinase. Do domain thụ thể L thứ nhất tham gia bám với phối tử IGF và domain Tyrosine kinase đóng vai trò quan trọng trong việc hoạt hóa tín hiệu sau khi IGF1R bám với các phối tử IGF, nên các exon quy định các domain này (exon 2 và các exon từ 16 đến 21) và các vùng trình tự intron xung quanh của gen *IGF1R* sẽ được khuếch đại để tìm kiếm các SNP. Các gen *IGFBP-1*, *-2*, *-3*, *-5*, *-6* của cá tra nuôi đều chứa 4 exon và 3 intron, exon thứ nhất chứa codon khởi đầu và exon thứ tư chứa codon kết thúc. Chiều dài các gen *IGFBP-1*, *-2*, *-3*, *-5*, *-6* lần lượt là 2,7 kb, 28,1 kb, 18,4 kb, 12,2 kb và 4,2 kb. Đầu N của các protein này đều chứa đoạn peptit tín hiệu và domain IGFBP được quy định bởi trình tự nằm gọn trong exon 1 của mỗi gen tương ứng. Đầu C của các protein này đều chứa domain Thyroglobulin type 1 được mã hóa bởi trình tự nằm trong exon 3 và 4 của các gen tương ứng. Để tìm kiếm SNP, các vùng gen thuộc vùng mã hóa cho các protein (CDS), vùng 5'-UTR, 3'-UTR cùng các vùng intron lân cận có thể liên quan đến điều hòa phiên mã của các gen này sẽ được lựa chọn để khuếch đại. Gen *IGFBP-7* của cá tra nuôi có chiều dài khoảng 9 kb với 5 exon, exon thứ nhất chứa codon khởi đầu và exon thứ năm chứa codon kết thúc. Đầu N của protein IGFBP-7 cũng chứa chuỗi peptit tín hiệu và domain IGFBP được quy định bởi trình tự nằm gọn trong exon 1 của gen. Khác với các protein IGFBP-1, *-2*, *-3*, *-5*, *-6*, protein IGFBP-7 lại chứa các domain KAZAL-FS, Ig3 và Ig ở đầu C, được mã hóa bởi các trình tự thuộc cả 5 exon. Do đó, trình tự của cả 5 exon và các vùng intron liền kề của gen *IGFBP-7* được khuếch đại để tìm kiếm SNP.

3.2. Xác định tính xác thực của trình tự giải mã bằng Sanger so với trình tự tham chiếu

Trình tự các đoạn gen được xác định trên một cá thể sau khi được giải mã bằng phương pháp Sanger được so sánh với trình tự đoạn gen tham chiếu tương ứng giải mã bằng NGS, thông qua phần mềm BLASTn, đã xác định độ tương đồng về trình tự nằm trong khoảng từ 99,9 đến 100 %, khẳng định được tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham chiếu, làm cơ sở để tiến hành nghiên cứu xác định đa hình trên gen đích theo hướng đã đề ra.

3.3. Phát hiện, sàng lọc SNP trên các gen đích thuộc hệ thống IGF trên bộ mẫu khởi tạo

Trình tự các đoạn gen đích của 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được giải mã bằng phương pháp Sanger và so sánh với trình tự tham chiếu để phát hiện SNP. Các tiêu chí sàng lọc SNP được trình bày ở mục 2.2.6.

3.3.1. Phát hiện và sàng lọc SNP trên gen *IGF1*

Bảng 3.1. Các SNP được phát hiện trên gen *IGF1*

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN ^(a)	TTC ^(a)	TTN	TTC
1	I1_13185 ⁽²⁾	C	T*	8CC:2CT (0,20)	6CC:2CT:2TT (0,50)	18C:2T	14C:6T
				p = 0,42		p = 0,23	
2	I1_13263 ⁽²⁾	T	C*	6TT:3TC:1CC (0,44)	7TT:3TC (0,30)	15T:5C	17T:3C
				p = 0,81		p = 0,69	
3	I1_13268	C	T*	10CC (0,00)	9CC:1CT (0,10)	20C	19C:1T
4	I1_13366	C	G*	9CC:1GG (0,11)	8CC:2CG (0,20)	18C:2G	18C:2G
5	I1_13367	C	A*	9CC:1AA (0,11)	9CC:1CA (0,10)	18C:2A	19C:1A
6	I1_13387	C	G*	8CC:2CG (0,20)	8CC:1CG:1GG (0,22)	18C:2G	17C:3G
7	I2_13680 ^(2,3,4)	A	T*	9AA:1TT (0,11)	4AA:1AT:5TT (1,50)	18A:2T	9A:11T
				p = 0,03		p = 0,005	
8	I2_13684 ⁽²⁾	G	C*	10GG (0,00)	7GG:1GC:2CC (0,38)	20G	16G:4C
				p = 0,11		p = 0,10	
9	I2_13820	C	T*	10CC (0,00)	9CC:1TT (0,11)	20C	18C:2T
10	I2_13843	T	A*	6TT:3TA:1AA	3TT:2TA:3AA:2NN	15T:5A	8T:8A

*: alen thiểu số, TTN: tăng trưởng nhanh, TTC: tăng trưởng chậm, ^(a): Tần số xuất hiện SNP của mỗi nhóm, ⁽²⁾, ⁽³⁾, ⁽⁴⁾: đáp ứng tiêu chí sàng lọc SNP số 2, 3, 4 tương ứng.

Mười SNP được phát hiện trên gen *IGF1*, tất cả đều nằm trên intron (Bảng 3.1). Bốn SNP 13185 C>T, 13263 T>C thuộc intron 1 và 13680 A>T và 13684 G>C thuộc intron 2 đạt tiêu chí có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên (Bảng 3.1). Tuy nhiên, chỉ có SNP 13680 A>T thể hiện sự khác biệt về thành phần kiểu gen và thành phần alen giữa hai nhóm, với giá trị p là 0,03 và 0,005 (Bảng 3.1) và được sàng lọc để kiểm nghiệm trên bộ mẫu lớn hơn.

3.3.2. Phát hiện và sàng lọc SNP trên gen *IGF2*

Mười hai SNP được phát hiện trên gen *IGF2*, trong đó, 3 SNP 1355 A>G, 1391 G>A ở intron 2 và 2061 T>A ở intron 3 có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên, nhưng không thể hiện được sự khác biệt có ý nghĩa về thành phần kiểu gen và/hoặc thành phần alen. Do đó không có SNP nào trên gen *IGF2* được sàng lọc để tiếp tục được kiểm nghiệm.

3.3.3. Phát hiện và sàng lọc SNP trên gen *IGF1R*

Bảng 3.3. Các SNP được phát hiện trên gen *IGF1R*

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN ^(a)	TTC ^(a)	TTN	TTC
1	II ₁₂₉₉₂ ⁽²⁾	A	C*	9AA:1AC (0,10)	6AA:2AC:2CC (0,50)	19A:1C	14A:6C
				p = 0,29		p = 0,09	
2	II ₁₃₃₅₇ ^(2,3)	T	C*	4TT:6TC (0,60)	8TC:2CC (1,25)	14T:6C	8T:12C
				p = 0,04		p = 0,11	
3	II ₁₅₃₅₉ ⁽²⁾	C	G*	7CC:3GG (0,43)	4CC:6GG (1,50)	14C:6G	8C:12G
				p = 0,37		p = 0,11	
4	II ₁₅₃₉₂ ^(2,3,4)	T	A*	5TT:5TA (0,50)	10TT (0,00)	15T:5A	20T
				p = 0,03		p = 0,04	
5	II ₁₅₃₉₇ ⁽²⁾	T	G*	5TT:3TG:2GG (0,63)	8TT:2TG (0,20)	13T:7G	18T:2G
				p = 0,23		p = 0,12	
6	II ₈₃₈₉₄ ^(2,4)	A*	G	4AA:1AG:5GG (0,83)	1AA:9GG (0,11)	9A:11G	2A:18G
				p = 0,08		p = 0,03	
7	II ₈₅₀₅₇ ⁽²⁾	T	C*	10TT (0,00)	7TT:1TC:2CC (0,38)	20T	16T:4C
				p = 0,11		p = 0,1	
8	II ₈₇₁₄₇ ⁽²⁾	T	C*	10TT (0,00)	7TT:3TC (0,30)	20T	17T:3C
				p = 0,21		p = 0,23	
9	II ₉₀₅₉₀ ⁽²⁾	A	T*	7AA:3AT (0,30)	10AA (0,00)	17A:3T	20A
				p = 0,21		p = 0,23	

*: alen thiếu số, TTN: tăng trưởng nhanh, TTC: tăng trưởng chậm, ^(a): Tần số xuất hiện SNP của mỗi nhóm, ⁽²⁾, ⁽³⁾, ⁽⁴⁾: đáp ứng tiêu chí sàng lọc SNP số 2, 3, 4 tương ứng.

Tất cả 9 SNP được phát hiện ở gen *IGF1R* nằm trong các intron và đều có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên (Bảng 3.3). Tuy nhiên, chỉ có 3 SNP thể hiện sự khác biệt về thành phần kiểu gen và/hoặc thành phần alen giữa hai nhóm là 13357 T>C, 15392 T>A và 83894 A>G, nên được sàng lọc để kiểm nghiệm trên bộ mẫu lớn hơn (Bảng 3.3).

3.3.4. Phát hiện và sàng lọc SNP trên gen *IGFBP-1*

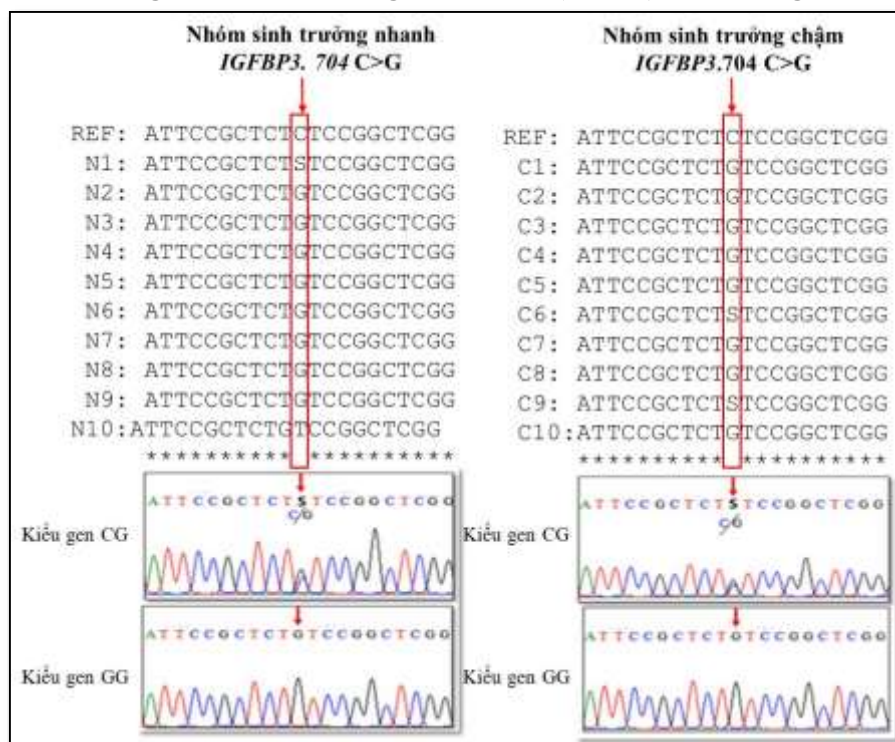
Bốn SNP được phát hiện trên gen này, trong đó có 1 SNP thuộc vùng promoter và 3 SNP thuộc các vùng intron. Tuy nhiên, không có SNP nào đạt đủ tiêu chí để được sàng lọc

3.3.5. Phát hiện và sàng lọc SNP trên gen *IGFBP-2*

Mười SNP được phát hiện trên gen *IGFBP-2* gồm 1 SNP thuộc vùng mã hóa, không làm thay đổi trình tự axit amin và 9 SNP nằm trong vùng không mã hóa, nhưng không có SNP nào đạt đủ tiêu chí để được sàng lọc.

3.3.6. Phát hiện và sàng lọc SNP trên gen *IGFBP-3*

Mười SNP được phát hiện, gồm 6 SNP thuộc các intron và 4 SNP thuộc các vùng 5'UTR, vùng mã hóa (CDS) và vùng 3'-UTR.



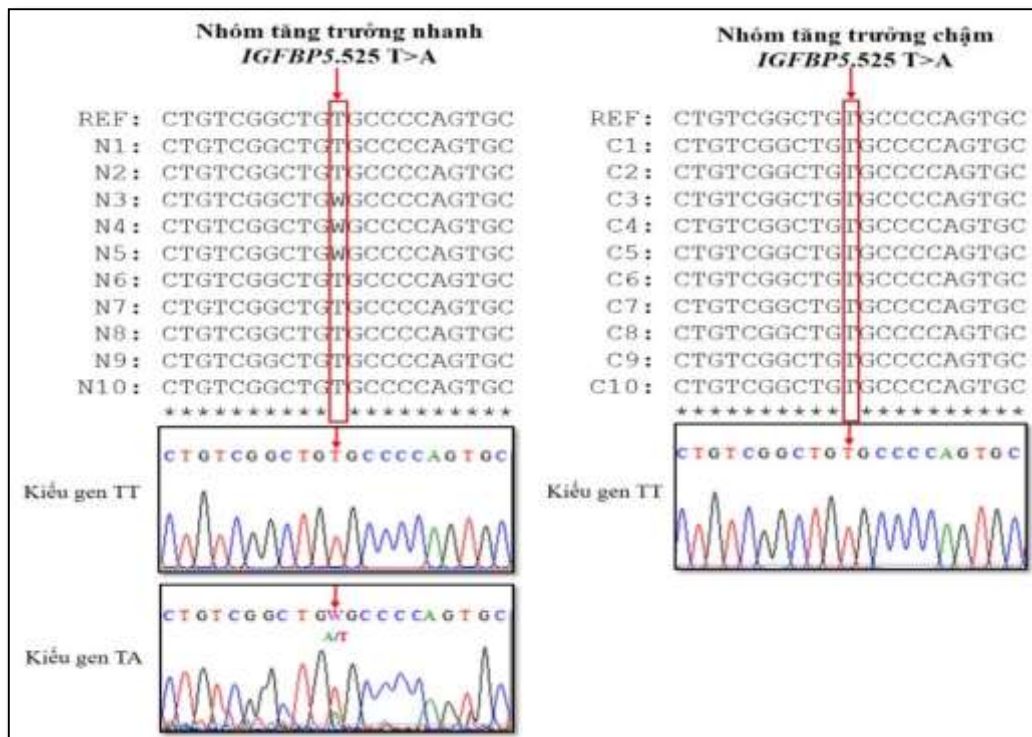
Hình 3.9: Kết quả phát hiện *IGFBP-3.704 C>G* trên gen *IGFBP-3* thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu.

REF: trình tự tham chiếu. N1 - N10: trình tự của 10 cá thể cá tra tăng trưởng nhanh. C1 - C10: trình tự của 10 cá thể cá tra tăng trưởng chậm. Mũi tên và khung màu đỏ: vị trí SNP được phát hiện trên mỗi cá thể.

SNP 704C>G ở exon 1 (Hình 3.9) nằm trong vùng mã hóa làm thay đổi trình tự axit amin thứ 8 Leucine thành Valine, nên SNP này được sàng lọc để kiểm nghiệm trên bộ mẫu lớn hơn. Các SNP còn lại nằm trong vùng không mã hóa không đạt đủ tiêu chí để được sàng lọc.

3.3.7. Phát hiện và sàng lọc SNP trên gen *IGFBP-5*

Bốn SNP được phát hiện trên gen thuộc các exon, trong đó, SNP 525T>A ở exon 1 gây ra sự thay đổi trình tự axit amin thứ 16 Valine thành axit Glutamic được sàng lọc để kiểm nghiệm lại trên bộ mẫu lớn hơn (Hình 3.10). Ba SNP còn lại không đạt đủ tiêu chí để được sàng lọc.



Hình 3.10. Kết quả phát hiện *IGFBP-5.525 T>A* trên gen *IGFBP-5* thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu.

REF: trình tự tham chiếu. N1 - N10: trình tự của 10 cá thể cá tra tăng trưởng nhanh. C1 - C10: trình tự của 10 cá thể cá tra tăng trưởng chậm. Mũi tên và khung màu đỏ: vị trí SNP được phát hiện trên mỗi cá thể.

3.3.8. Phát hiện và sàng lọc SNP trên gen *IGFBP-6*

Mười bảy SNP được phát hiện ở gen *IGFBP-6*, đều thuộc vùng không mã hóa. Trong đó, có 7 SNP có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên gồm: 560 A>G, 2275 C>A, 2278 C>A, 2423 G>T, 3177 T>C, 3469 T>C và 3505 G>A (Bảng 3.8, trích). Tuy nhiên, chỉ có SNP 2278 C>A được sàng lọc do có sự khác biệt có ý nghĩa về thành phần kiểu gen ($p = 0,02$) (Bảng 3.8, trích).

Bảng 3.8. Các SNP được phát hiện trên gen *IGFBP-6* (trích)

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN ^(a)	TTC ^(a)	TTN	TTC
1	I1_560 ⁽²⁾	A	G*	6AA:2AG:2GG (0,50)	8AA:1AG:1GG (0,22)	14A:6G	17A:3G
				p = 0,52		p = 0,45	
11	I3_2275 ⁽²⁾	C	A*	8CC:2AA (0,25)	5CC:5AA (1,00)	16C:4A	10C:10A
				p = 0,35		p = 0,09	
12	I3_2278 ^(2,4)	C	A*	8CC:2AA (0,25)	4CC:6AA (1,50)	16C:4A	8C:12A
				p = 0,17		p = 0,02	
13	I3_2423 ⁽²⁾	G	T*	6GG:4TT (0,67)	4GG:6TT (1,50)	12G:8T	8G:12T
				p = 0,66		p = 0,34	
15	E4 3'UTR 3177 ⁽²⁾	T	C*	6TT:4CC (0,67)	6TT:4CC (0,67)	12T:8C	12T:8C
				p = 1,0		p = 1,0	
16	E4 3'UTR 3469 ⁽²⁾	T*	C	3TT:2TC:5CC (0,71)	1TT:9CC (0,11)	8T:12C	2T:18C
				p = 0,13		p = 0,06	
17	E4 3'UTR 3505 ⁽²⁾	G	A*	7GG:3AA (0,43)	6GG:4AA (0,67)	14G:6A	12G:8A
				p = 1,0		p = 0,74	

*: alen thiểu số, TTN: tăng trưởng nhanh, TTC: tăng trưởng chậm, ^(a): Tần số xuất hiện SNP của mỗi nhóm, ⁽²⁾, ⁽⁴⁾: đáp ứng tiêu chí sàng lọc SNP số 2, 4 tương ứng.

3.3.9. Phát hiện và sàng lọc SNP trên gen *IGFBP-7*

Bảng 3.9. Các SNP được phát hiện trên gen *IGFBP-7* (trích)

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN ^(a)	TTC ^(a)	TTN	TTC
2	E1 - CDS 344 ⁽¹⁾	T*	C	1TT:2TC:7CC	3TT:3TC:4CC	4T:16C	9T:11C
12	I2 - 2060 ^(2,3,4)	A	G*	10GG (0,00)	3AA:2AG:5GG (1,40)	20G	8A:12G
				p = 0,02		p = 0,003	
20	E3 - CDS 4559 ⁽¹⁾	C*	A	2CC:3CA:5AA	10AA	7C:13A	20A

*: alen thiểu số, TTN: tăng trưởng nhanh, TTC: tăng trưởng chậm, ^(a): Tần số xuất hiện SNP của mỗi nhóm, ⁽¹⁾, ⁽²⁾, ⁽³⁾, ⁽⁴⁾: đáp ứng tiêu chí sàng lọc SNP số 1, 2, 3, 4 tương ứng.

Có 27 SNP được phát hiện trên gen *IGFBP-7*, trong đó 2 SNP gây sự thay đổi axit amin là 344 T>C (p.Leu78Pro) và 4559 C>A (p.Leu189Met) (Bảng 3.9, trích). Trong các SNP còn lại, 17 SNP có tần số xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên, bao gồm -376 T>G ở promoter, 561 G>A, 598 C>A, 1052 T>A, 1117 C>G, 1160 G>C và 1216 C>A ở intron 1, 2060 A>G, 2138 T>C, 3981 T>G, 4058 A>G, 4334 G>A và 4465 C>T ở intron 2, 4685 T>C ở intron 3, 7405 C>T, 7676 G>C và 7692 T>C ở

intron 4. Tuy nhiên, chỉ có SNP 2060 A>G đạt tiêu chí về sự khác biệt có ý nghĩa về thành phần kiểu gen/thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm (với cả 2 giá trị p nhỏ hơn 0,05) (Bảng 3.9, trích). Do vậy, 3 SNP 344 T>C (p.Leu78Pro), 2060 A>G và 4559 C>A (p.Leu189Met) được sàng lọc để tiếp tục genotyping ở bộ mẫu lớn hơn.

3.4. Phân tích sự liên quan đến tính trạng tăng trưởng của các SNP được sàng lọc

3.4.1. SNP genotyping trên các cá thể thuộc bộ mẫu kiểm nghiệm

Các đoạn gen chứa 10 SNP được sàng lọc ở các gen *IGF1*, *IGF1R*, *IGFBP-3*, -5, -6, -7 của 140 cá thể thuộc bộ mẫu kiểm nghiệm được phân lập thành công, làm khuôn cho phản ứng SBE xác định SNP. Dữ liệu SNP thu được từ tổng cộng 160 cá thể của bộ mẫu khởi tạo và bộ mẫu kiểm nghiệm được gộp lại để phân tích sự liên quan của các SNP được sàng lọc và tính trạng tăng trưởng.

3.4.2. Phân tích mối liên quan giữa SNP được sàng lọc và tính trạng tăng trưởng của cá tra nuôi

3.4.2.1. SNP được sàng lọc trên gen *IGF1*

Sự khác biệt về thành phần kiểu gen phát sinh từ SNP *IGF1*.13680 A>T giữa hai nhóm tăng trưởng nhanh/chậm có ý nghĩa với p bằng 0,009 (Bảng 3.11), cho thấy SNP này có liên quan đến tính trạng tăng trưởng.

Bảng 3.11. Phân tích mối liên quan giữa SNP 13680 A>T trên gen *IGF1* với tính trạng tăng trưởng của cá tra nuôi

Các thông số	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p
Thành phần kiểu gen (^b)	AA (18) AT (61) TT (01)	AA (07) AT (64) TT (07) NN (02)	0,009**
Thành phần alen (^c)	A (97) T (63)	A (78) T (78)	0,06
PIC	0,494		
MAF	0,446		

(^b): Số lượng cá thể mang kiểu gen tương ứng, NN: cá thể không xác định được kiểu gen, (^c): số lượng alen tương ứng, **: p < 0,01.

3.4.2.2. Các SNP được sàng lọc trên gen *IGF1R*

Trong số 3 SNP được sàng lọc trên gen *IGF1R* 13357 T>C, 15392T>A và 83894 A>G, chỉ còn SNP 13357 T>C thể hiện được sự khác biệt về thành phần kiểu gen giữa nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm (giá trị p bằng 0,03) (Bảng 3.12).

Bảng 3.12. Phân tích mối liên quan giữa 3 SNP trên gen *IGF1R* với tính trạng tăng trưởng của cá tra nuôi.

Tên SNP	Thành phần kiểu gen ^(b)			Thành phần alen ^(c)			PIC	MAF
	TTN	TTC	p	TTN	TTC	p		
13357 T>C	TT (32) TC (47) CC (01)	TT (26) TC (44) CC (09) NN (01)	0,03*	T (111) C (49)	T (96) C (62)	0,11	0,454	0,349
15392 T>A	TT (06) TA (68) NN (06)	TT (10) TA (64) AA (01) NN (05)	0,35	T (80) A (68)	T (84) A (66)	0,74	0,495	0,450
83894 A>G	AA (09) AG (66) GG (05)	AA (02) AG (69) GG (09)	0,06	A (84) G (76)	A (73) G (87)	0,22	0,499	0,491

TTN: Tăng trưởng nhanh, TTC: Tăng trưởng chậm, ^(b): Số lượng cá thể mang kiểu gen tương ứng, NN: cá thể không xác định được kiểu gen, ^(c): số lượng alen tương ứng, *: $p < 0,05$.

Kết quả phân tích haplotype cho thấy haplotype *IGF1R_H1* (CAA), H4 (TTA) và H5 (TTG) có mối liên quan chặt chẽ với nhóm tăng trưởng nhanh ($p < 0,01$), trong khi haplotype H2 (CTG), H3 (TAA) có liên quan chặt chẽ với nhóm tăng trưởng chậm ($p < 0,01$) (Bảng 3.14).

Bảng 3.14. Phân tích mối liên quan giữa haplotype phát sinh từ 3 SNP của gen *IGF1R* lên tính trạng tăng trưởng của cá tra nuôi.

Haplotype tạo nên từ 3 SNP				TTN ^(d)	TTC ^(d)	P
13357 T>C	15392 T>A	83894 A>G	Kí hiệu			
C	A	A	<i>IGF1R_H1</i>	35,39 (0,239)	10,18 (0,068)	2,73e-5***
C	T	G	<i>IGF1R_H2</i>	3,77 (0,025)	46,79 (0,312)	7,47e-11***
T	A	A	<i>IGF1R_H3</i>	28,96 (0,196)	54,78 (0,365)	0,001**
T	T	A	<i>IGF1R_H4</i>	11,15 (0,075)	1,01(0,007)	0,002**
T	T	G	<i>IGF1R_H5</i>	62,59 (0,423)	34,17 (0,228)	0,00018***

TTN: Tăng trưởng nhanh, TTC: Tăng trưởng chậm, ^(d): tần số xuất hiện haplotype trong mỗi nhóm, haplotype có tần số xuất hiện trong cả hai nhóm đều nhỏ hơn 0,03 sẽ không được phân tích giá trị p, ** $p < 0,01$, *** $p < 0,001$.

3.4.2.3. SNP được sàng lọc trên gen *IGFBP-3*

Thành phần kiểu gen phát sinh từ SNP 704 C>G (Leu8Val) giữa hai nhóm tăng trưởng nhanh/chậm khác biệt có ý nghĩa, với giá trị p là 0,03 ($p < 0,05$) (Bảng 3.15). Ngoài ra, thành phần alen giữa hai nhóm cũng khác biệt có ý nghĩa thống kê, với giá trị p là 0,013 ($p < 0,05$) (Bảng 3.15).

Bảng 3.15. Phân tích mối liên quan giữa SNP 704 C>G (p.Leu8Val) trên gen *IGFBP-3* với tính trạng tăng trưởng của cá tra nuôi

Các thông số	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p
Thành phần kiểu gen ^(b)	CG (45) GG (35)	CG (26) GG (53) NN (01)	0,03*
Thành phần alen ^(c)	C (45) G (115)	C (26) G (132)	0,013*
PIC	0,347		
MAF	0,223		

^(b): Số lượng cá thể mang kiểu gen tương ứng, NN: cá thể không xác định được kiểu gen, ^(c): số lượng alen tương ứng, * $p < 0,05$.

3.4.2.4. SNP được sàng lọc trên gen *IGFBP-5*

Thành phần kiểu gen và thành phần alen phát sinh từ SNP 525 T>A (p.Val16Glu) khác biệt có ý nghĩa giữa hai nhóm tăng trưởng nhanh/chậm (với cả 2 giá trị p đều nhỏ hơn 0,001) (Bảng 3.16).

Bảng 3.16. Phân tích mối liên quan giữa SNP 525 T>A (p.Val16Glu) trên gen *IGFBP-5* với tính trạng tăng trưởng của cá tra nuôi

Các thông số	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p
Thành phần kiểu gen ^(b)	TT (07) TA (69) AA (03) NN (01)	TT (41) TA (37) NN (02)	1,13e-8***
Thành phần alen ^(c)	T (83) A (75)	T (119) A (37)	1,14e-5***
PIC	0,459		
MAF	0,357		

^(b): Số lượng cá thể mang kiểu gen tương ứng, NN: cá thể không xác định được kiểu gen, ^(c): số lượng alen tương ứng, * $p < 0,05$, ** $p < 0,01$.

3.4.2.5. SNP được sàng lọc trên gen *IGFBP-6*

SNP 2278 C>A trên gen *IGFBP-6* sau khi được kiểm nghiệm không cho thấy sự khác biệt có ý nghĩa giữa thành phần kiểu gen hay thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm.

3.4.2.6. Các SNP được sàng lọc trên gen *IGFBP-7*

Kết quả kiểm nghiệm cho thấy SNP 344 T>C (p.Leu78Pro) không thể hiện được sự khác nhau về thành phần kiểu gen/thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, với các giá trị p đều lớn hơn 0,05. SNP 2060 A>G thể hiện sự khác biệt có ý nghĩa về cả thành phần kiểu gen/thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, với cả hai giá trị p đều nhỏ hơn 0,001. SNP 4559 C>A (p.Leu189Met) vẫn thể hiện được sự khác biệt về thành phần kiểu gen giữa hai nhóm với p nhỏ hơn 0,05

(Bảng 3.18). Như vậy, 2 SNP 2060 A>G và 4559 C>A (p.Leu189Met) được coi là 2 chỉ thị SNP tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

Bảng 3.18. Phân tích mối liên quan giữa 3 SNP trên gen *IGFBP-7* với tính trạng tăng trưởng của cá tra nuôi.

Tên SNP	Thành phần kiểu gen ^(b)			Thành phần alen ^(c)			PIC	MAF
	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p		
344 T>C	TT (01) TC (56) CC (23)	TT (03) TC (44) CC (31) NN (02)	0,165	T (58) C (102)	T (50) C (106)	0,43	0,449	0,342
2060 A>G	AA (04) AG (46) GG (30)	AA (35) AG (37) GG (06) NN (02)	1,01e-9***	A (54) G (106)	A (107) G (49)	6,34e-10***	0,499	0,491
4559 C>A	CC (04) CA (07) AA (69)	CC (03) CA (18) AA (57)	0,047*	C (15) A (145)	C (24) A (132)	0,104	0,216	0,123

^(b): Số lượng cá thể mang kiểu gen tương ứng, NN: cá thể không xác định được kiểu gen, ^(c): số lượng alen tương ứng, *: $p < 0,05$ **: $p < 0,01$, ***: $p < 0,001$.

Tạo nên từ 3 SNP được sàng lọc trên gen *IGFBP-7*, haplotype *IGFBP-7_H1* (CAA) và H2 (CAC) có mối liên quan với nhóm tăng trưởng chậm với giá trị p nhỏ hơn 0,001. Trong khi đó, haplotype *IGFBP-7_H3* (CGA) và H4 (CGC) tương quan chặt chẽ với nhóm tăng trưởng nhanh với giá trị p nhỏ hơn 0,001 và p nhỏ hơn 0,01, tương ứng (Bảng 3.20).

Bảng 3.20. Phân tích mối liên quan giữa haplotype phát sinh từ 3 SNP của gen *IGFBP-7* lên tính trạng tăng trưởng

Haplotype tạo nên từ 3 SNP				TTN ^(d)	TTC ^(d)	p
344 T>C	2060 A>G	4559 C>A	Ký hiệu			
C	A	A	<i>IGFBP7_H1</i>	20,32 (0,127)	58,84 (0,377)	2,97e-7***
C	A	C	<i>IGFBP7_H2</i>	2,75 (0,017)	22,13 (0,142)	3,97e-5***
C	G	A	<i>IGFBP7_H3</i>	66,69 (0,417)	23,17 (0,148)	1,31e-7***
C	G	C	<i>IGFBP7_H4</i>	12,25 (0,077)	1,86 (0,012)	0,0054**
T	A	A	<i>IGFBP7_H5</i>	30,93 (0,193)	26,02 (0,167)	0,54
T	G	A	<i>IGFBP7_H6</i>	27,07 (0,169)	23,97 (0,154)	0,708

TTN: Tăng trưởng nhanh, TTC: Tăng trưởng chậm, ^(d): tần số xuất hiện haplotype trong mỗi nhóm, haplotype có tần số xuất hiện trong cả hai nhóm đều nhỏ hơn 0,03 sẽ không được phân tích giá trị p, ** $p < 0,01$, *** $p < 0,001$.

3.4.2.5. Tác động tổng hợp của các SNP có tiềm năng liên quan đến tăng trưởng

Bảng 3.21. Phân tích tác động của các haplotype phát sinh từ 6 chỉ thị SNP được lựa chọn lên tính trạng tăng trưởng của cá tra nuôi

Haplotype tạo thành từ 6 SNP						Ký hiệu	TTN ^(d)	TTC ^(d)	p
<i>IGF1</i> 13680 A>T	<i>IGF1R</i> 13357 T>C	<i>IGF</i> <i>BP3</i> 704 C>G	<i>IGF</i> <i>BP5</i> 525 T>A	<i>IGF</i> <i>BP7</i> 2060 A>G	<i>IGF</i> <i>BP7</i> 4559 C>A				
A	T	G	T	A	A	CO-H1	5,39 (0,034)	26,92 (0,179)	7,18e-5 ***
A	T	G	T	G	A	CO-H2	65,44 (0,414)	4,69 (0,031)	2,22e-16 ***
A	C	G	T	A	A	CO-H3	0,00 (0,000)	33,54 (0,224)	9,09e-10 ***
T	T	C	A	A	A	CO-H4	6,98 (0,044)	12,46 (0,083)	0,218
T	T	C	A	G	A	CO-H5	9,37 (0,059)	1,99 (0,013)	0,022 *
T	T	G	A	A	A	CO-H6	6,34 (0,040)	2,32 (0,015)	0,152
T	T	G	T	A	A	CO-H7	0,00 (0,000)	7,88 (0,053)	0,004 **
T	T	G	T	A	C	CO-H8	0,00 (0,000)	8,52 (0,057)	0,003 **
T	T	G	T	G	A	CO-H9	1,20 (0,008)	7,70 (0,051)	0,029 *
T	T	G	A	G	A	CO-H10	2,27 (0,014)	9,64 (0,064)	0,032 *
T	C	C	A	A	A	CO-H11	16,54 (0,105)	0,00 (0,000)	2,04e-5 ***
T	C	C	A	G	A	CO-H12	3,20 (0,020)	4,51 (0,030)	0,658
T	C	G	T	G	A	CO-H13	0,00 (0,000)	7,79 (0,052)	0,005 **
T	C	G	A	A	A	CO-H14	7,84 (0,050)	0,00 (0,000)	0,003 **

CO: Combination, TTN: Tăng trưởng nhanh, TTC: Tăng trưởng chậm, ^(d): tần số xuất hiện haplotype trong mỗi nhóm, * $p < 0,05$, ** $p < 0,01$.

Tác động tổng hợp của 6 SNP tiềm năng liên quan đến tính trạng tăng trưởng gồm: *IGF1*. 13680 A>T, *IGF1R*. 13357 T>C, *IGFBP-3*. 704C>G (p.Leu8Val), *IGFBP-5*. 525T>A (p.Val16Glu), *IGFBP-7*. 2060 A>G và *IGFBP-7*. 4559 C>A (p.Leu189Met) được phân tích thông qua sự khác biệt về tỉ lệ haplotype/tổ hợp kiểu gen giữa hai nhóm tăng trưởng nhanh/chậm, cho thấy 4 haplotype chiếm tỉ lệ cao hơn ở nhóm tăng trưởng nhanh là CO_H2, H5, H11, H14 và 7 haplotype chiếm tỉ lệ cao hơn ở

nhóm tăng trưởng chậm là CO_H1, H3, H7, H8, H9, H10, H14 (Bảng 3.21).

Bảng 3.22. Phân tích tác động của các tổ hợp kiểu gen phát sinh từ 6 chỉ thị SNP được lựa chọn lên tính trạng tăng trưởng của cá tra nuôi

Tổ hợp kiểu gen được tạo thành từ 6 SNP						TTN (^e)	TTC (^e)	p
<i>IGF1</i> 13680 A>T	<i>IGF1R</i> 13357 T>C	<i>IGF</i> <i>BP3</i> 704 C>G	<i>IGF</i> <i>BP5</i> 525 T>A	<i>IGF</i> <i>BP7</i> 2060 A>G	<i>IGF</i> <i>BP7</i> 4559 C>A			
AA	CT	GG	AT	GG	AA	3 (3,75%)	0 (0%)	0,083
AA	CT	GC	AT	GA	AA	3 (3,75%)	0 (0%)	0,083
AT	CT	GG	AT	GG	AA	3 (3,75%)	0 (0%)	0,083
AT	CT	GG	AT	GA	AA	8 (10%)	6 (7,5%)	0,592
AT	CT	GG	TT	AA	AA	0 (0%)	3 (3,75%)	0,083
AT	CT	GG	TT	AA	AC	0 (0%)	3 (3,75%)	0,083
AT	CT	GG	TT	AG	AA	0 (0%)	3 (3,75%)	0,083
AT	CT	GC	AT	GG	AA	3 (3,75%)	0 (0%)	0,083
AT	CT	GC	AT	GA	AA	14 (17,5%)	5 (6,25%)	0,038*
AT	CT	GC	AT	AA	AA	2 (2,5%)	6 (7,5%)	0,157
AT	TT	GG	AT	GA	AA	4 (5%)	4 (5%)	1,000
AT	TT	GG	TT	AA	AC	0 (0%)	3 (3,75%)	0,083
AT	TT	GG	TT	AG	AA	0 (0%)	3 (3,75%)	0,083
AT	TT	GG	TT	AG	AC	0 (0%)	3 (3,75%)	0,083
AT	TT	GC	AT	GG	AA	8 (10%)	0 (0%)	0,004**
AT	TT	GC	AT	GA	AA	6 (7,5%)	1 (1,25%)	0,058
AT	TT	GC	AT	AA	AA	1 (1,25%)	4 (5%)	0,179
TT	CT	GG	TT	GG	AA	0 (0%)	3 (3,75%)	0,083

CO; Combination, (^e): tần số tổ hợp kiểu gen trong mỗi nhóm, tổ hợp kiểu gen có tần số xuất hiện trong cả hai nhóm đều nhỏ hơn 3% sẽ không được phân tích giá trị p, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

Tạo nên từ 6 SNP tiềm năng theo thứ tự nêu trên, tổ hợp kiểu gen AT, CT, GC, AT, GA và AA (tạo nên từ haplotype CO_H2 và CO_H11 (diplotype H2/H11)) có tần số cao hơn ở nhóm tăng trưởng nhanh, tổ hợp kiểu gen AT, TT, GC, AT, GG và AA (tạo nên từ haplotype CO_H2 và CO_H5 (diplotype H2/H5)) chỉ xuất hiện ở nhóm tăng trưởng nhanh (các giá trị $p < 0,05$) (Bảng 3.22).

CHƯƠNG 4. THẢO LUẬN

4.1. Trình tự và cấu trúc hoàn chỉnh của các gen thuộc hệ thống IGF của cá tra nuôi được phân tích tương đồng với các loài cá xương khác

Các gen đích thuộc hệ thống IGF của cá tra nuôi được phân tích đều có mức độ tương đồng về trình tự cao nhất với các trình tự tương ứng của cá nheo Mỹ *Ictalurus punctatus* rất gần với cá tra nuôi trong hệ thống phân loại, thuộc bộ cá da trơn *Siluriformes*. Cấu trúc gen/protein IGF1, IGF1R, IGFBP-1, -2, -3, -5, -6, -7 của cá tra nuôi tương tự với cấu trúc gen, protein tương ứng ở các loài cá xương khác, với các domain chức năng điển hình.

4.2. Sự đa dạng di truyền của các gen thuộc hệ thống IGF của cá tra nuôi

Trong luận án, số lượng SNP nằm trong vùng không mã hóa chiếm tỉ lệ 93,2%, phù hợp với thống kê từ các nghiên cứu về đa hình ở quy mô hệ gen đã khẳng định tỉ lệ này khoảng 90%. Ở quy mô gen đích, các nghiên cứu tìm kiếm SNP trên các gen liên quan đến tăng trưởng của cá như: gen Growth Hormone (*GH*) trên cá quế *Siniperca chuatsi*, gen *IGF1* trên cá chép *C. carpio*, gen *myostatin-1* trên cá ngão *Ancherythroculter nigrocauda*, gen *GH*, *IGF1* và *Myogenin (MyoG)* trên cá rô phi sông Nin *O. niloticus* cũng đưa tới nhận định này. Ngược lại, số lượng SNP gây thay đổi trình tự axit amin được phát hiện trong luận án chỉ chiếm 3,8% tổng số SNP được phát hiện, đây là một tỉ lệ nhỏ, có thể là do các axit amin bị thay đổi sẽ dẫn đến sự hạn chế về chức năng, nên các đa hình dạng này không được ưu tiên giữ lại trong quá trình tiến hóa. Điều này phù hợp với một số nghiên cứu khác trên cá xương cũng khẳng định tỉ lệ các đa hình gây sự thay thế axit amin trên tổng số SNP được phát hiện trên các gen chiếm tỉ lệ thấp, khoảng 4% ở cá hồi *S. salar*, 10% ở cá nheo Mỹ *I. punctatus*, cá bơn *Solea solea L.*, 11% đến 14% ở cá bơn Đại Tây Dương *Scophthalmus maximus*.

Chỉ có 10 SNP được sàng lọc thuộc các gen *IGF1*, *IGF1R*, *IGFBP-3*, -5, -6 và *IGFBP-7* để genotyping với bộ mẫu kiểm nghiệm, chiếm tỉ lệ 9,7%. Việc khu trú được một phần nhỏ các SNP có khả năng cao liên quan đến tính trạng thông qua các tiêu chí sàng lọc sơ bộ, để tiếp tục phân tích ở các bước tiếp theo là rất cần thiết để đảm bảo tính chính xác mà vẫn tiết kiệm thời gian và chi phí nghiên cứu. Hướng nghiên cứu này tương tự nghiên cứu trên cá hồi Đại Tây Dương *S. salar*, sử dụng 60 trên 7000 SNP

để xác định cá hồi tự nhiên và cá hồi nuôi, trên cá nheo lục *Ictalurus furcatus* khi sử dụng 64 trên 4275 SNP để genotyping xác định cá tự nhiên và cá được gia hóa, trên cá quế *S. chuatsi* khi sử dụng 48 trên 5205 SNP để đánh giá các tính trạng tăng trưởng.

Không có SNP được sàng lọc trên các gen *IGF2*, *IGFBP-1* và gen *IGFBP-2* của cá tra nuôi. Các nghiên cứu trên thế giới phần lớn mới chỉ ra được sự liên quan ở mức độ biểu hiện của ba gen này với các tính trạng tăng trưởng cá nheo Mỹ *I. punctatus*, cá lưỡi trâu *Cynoglossus semilaevis*, cá rô phi sông Nin *O. niloticus*, cá vược châu Âu *D. labrax*, các loài cá thuộc họ Cá chép, cá ngựa vằn *Danio rerio*, nhưng có rất ít hoặc chưa có nghiên cứu chỉ ra mối liên quan chặt chẽ giữa SNP trên các gen này với các tính trạng tăng trưởng của cá xương.

Bốn SNP *IGFBP-3*. 704C>G, *IGFBP-5*. 525T>A, *IGFBP-7*. 344 T>C và *IGFBP-7*. 4559 C>A gây thay đổi trình tự axit amin Leu8Val, Val16Glu, Leu78Pro và Leu189Met trên protein tương ứng. Trong đó, Leu8Val, Val16Glu được xác định nằm trong chuỗi peptit tín hiệu ở đầu N của protein *IGFBP-3* và *IGFBP-5*. Do chuỗi peptit tín hiệu đóng một vai trò quan trọng trong việc vận chuyển và bài tiết protein *IGFBP-3* và *IGFBP-5*, nên tác động của những sự thay đổi trình tự axit amin trên lên chức năng của hai protein này của cá tra nuôi sẽ cần tiếp tục được nghiên cứu sâu hơn. So sánh trình tự protein *IGFBP-7* của cá tra nuôi và một số loài cá xương khác cho thấy Leucine ở vị trí 78, bị thay thế bởi Proline do SNP 344T>C, không nằm trong vùng bảo thủ, trong khi Leucine ở vị trí 189, bị thay thế bởi Methionine do SNP 4559C>A, lại nằm trong vùng bảo thủ thuộc domain Ig và Ig3, vốn có chức năng chính trong miễn dịch.

4.3. Sự liên quan giữa đa dạng di truyền của các gen thuộc hệ thống IGF với tính trạng tăng trưởng của cá tra nuôi

SNP 13680 A>T ở intron 2 của gen *IGF1* thể hiện sự liên quan đến tính trạng tăng trưởng của cá tra nuôi, phù hợp với các nghiên cứu khác cũng tìm ra các SNP nằm trên vùng không mã hóa của gen *IGF1* có liên quan đến tính trạng tăng trưởng của các loài cá xương như cá chép *C. carpio*, cá hồi Đại Tây Dương *S. salar* và trên các loài gia súc và chim. Trong số 3 SNP được sàng lọc và genotyping của gen *IGF1R* của cá tra nuôi, chỉ có SNP 13357 T>C trên intron 1 tiếp tục thể hiện được sự khác biệt về thành phần kiểu gen giữa hai nhóm tăng trưởng nhanh/chậm, tương tự nghiên cứu về đa hình trên gen *IGF1R* trên một loài cá bản địa Trung

Quốc *O. potamophila* cũng xác định được 3 SNP trên vùng không mã hóa nhưng chỉ có 1 SNP liên quan đến các tính trạng tăng trưởng quan trọng. Phần lớn các SNP trên gen *IGFBP-3* và *IGFBP-5* liên quan đến tính trạng tăng trưởng của người, lợn, cá chép, gà đều nằm trong vùng không mã hóa. Như vậy, luận án đã có phát hiện mới về 2 SNP nằm trên vùng mã hóa làm thay đổi trình tự axit amin là *IGFBP-3*. 704 C>G (p.Leu8Val) và *IGFBP-5*. 525 T>A (p.Val16Glu) có liên quan đến sự tăng trưởng của cá tra nuôi. Điều này, mặt khác lại phù hợp với các nghiên cứu về mối liên quan giữa các SNP trên các gen *FOXO*, *MSTN*, *Myogenin* gây sự thay thế axit amin với tính trạng tăng trưởng của các loài thủy sản như ngao, *Culter alburnus* (♀) x *Ancherythroculter nigrocauda* (♂) và cá nheo Mỹ *I. punctatus*. Có một SNP ở trên gen *IGFBP-6* được sàng lọc để kiểm nghiệm nhưng không thể hiện sự liên quan tới tính trạng tăng trưởng của cá tra nuôi. Điều này cũng phù hợp với tình hình nghiên cứu hiện chưa sáng tỏ về vai trò và chức năng của *IGFBP-6* đối với sự tăng trưởng của cá xương. Hai SNP 2060 A>G trên intron 2 và 4559 C>A (p.Leu189Met) trên gen *IGFBP-7* thể hiện được mối liên quan đến tính trạng tăng trưởng của cá tra nuôi, củng cố thêm mối liên quan chặt chẽ giữa tính trạng tăng trưởng của cá xương với các SNP trên intron, cũng như với các SNP làm thay đổi trình tự axit amin như các nghiên cứu trước đó đã chỉ ra. Sáu SNP tiềm năng liên quan đến tăng trưởng có các giá trị PIC đa phần nằm trong khoảng 0,25 đến 0,5, cho thấy mức độ đa dạng di truyền trung bình, là cơ sở tiềm năng cho việc sử dụng chúng trong chọn giống; và phần lớn các giá trị MAF trong khoảng từ 20 đến 50 %, cho thấy các SNP này là các biến dị thường gặp, góp phần đáng kể vào sự đa dạng di truyền. Riêng giá trị PIC và MAF của SNP *IGFBP-7*. 4559C>A (p.Leu189Met) thấp hơn so với các SNP tiềm năng còn lại và không nằm trong khoảng giá trị trên, có thể do SNP này gây thay đổi trình tự axit amin trong vùng bảo thủ của domain Ig-Ig3 của protein *IGFBP-7*.

Đối với gen *IGF1R*, mặc dù SNP 15392 T>A và 83894 A>G không thể hiện được mối liên quan đến tính trạng tăng trưởng của cá tra nuôi khi đứng riêng lẻ nhưng các haplotype được tạo nên từ 2 SNP này và SNP 13357 T>C trên cùng gen đó lại liên quan chặt chẽ đến tính trạng tăng trưởng. Tương tự, đối với gen *IGFBP-7*, SNP 344 T>C (p.Leu78Pro) không liên quan đến tính trạng tăng trưởng của cá tra nuôi, nhưng sự kết hợp của SNP này với 2 chỉ thị SNP tiềm năng trên cùng gen là 2060 A>G

và 4559 C>A (p.Leu189Met) cũng chỉ ra các haplotype ưu thế ở mỗi nhóm tăng trưởng. Kết quả này chỉ ra rằng các SNP đứng riêng lẻ đôi khi không có ý nghĩa, nhưng khi được phân tích kết hợp với các SNP khác dưới dạng haplotype lại thể hiện được sự liên quan đến tính trạng quan tâm, tương tự như với tính trạng tăng trưởng của cá vược miệng rộng *Micropterus salmoides*, cá nheo Mỹ *I. punctatus* và nhấn mạnh vai trò của chỉ thị haplotype với tính trạng tăng trưởng như nghiên cứu trên cá hồi Đại Tây Dương *S. salar*.

4.4. Phát triển chỉ thị phân tử SNP nhằm ứng dụng chọn giống cá tra theo hướng tăng trưởng

Tác động tổng hợp của 6 SNP tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi được phân tích thông qua sự khác biệt về tần số haplotype cũng như tổ hợp kiểu gen giữa hai nhóm tăng trưởng nhanh/chậm đã cho thấy sự thống nhất, hợp lý và xuyên suốt của các kết quả thu được, cung cấp thêm các chỉ thị haplotype, tổ hợp kiểu gen tiềm năng liên quan đến tính trạng quan tâm. Từ đó, tính trạng tăng trưởng của cá tra nuôi cũng được nhận định là tính trạng đa gen, tương tự như tính trạng tăng trưởng của cá hồi vân *Oncorhynchus mykiss*, cá chép *C. carpio*, củng cố cho việc phát triển chỉ thị phân tử dạng panel, tích hợp nhiều SNP tiềm năng trên các gen liên quan khác nhau để đánh giá tính trạng tăng trưởng.

KẾT LUẬN VÀ KIẾN NGHỊ

Kết luận

1. Trình tự và cấu trúc của 9 gen thuộc hệ thống IGF (bao gồm *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7) đã được phân tích, các vùng gen chứa SNP liên quan đến tính trạng tăng trưởng của cá tra nuôi được xác định và giải mã.
2. Tổng số 103 SNP trên 9 gen (*IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6 và -7) của các cá thể cá tra nuôi thuộc bộ mẫu khởi tạo gồm cá thể tăng trưởng nhanh và cá thể tăng trưởng chậm được phát hiện, 10/103 SNP này được chọn lọc cho kiểm nghiệm trên quần thể lớn.
3. Các SNP được kiểm nghiệm trên quần thể lớn bằng phương pháp kéo dài một nucleotide (Single base extension) và thu được 6 SNP tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi bao gồm: *IGF1*. 13680 A>T, *IGF1R*. 13357 T>C, *IGFBP-3*. 704C>G (p.Leu8Val), *IGFBP-5*.

525T>A (p.Val16Glu), *IGFBP-7*. 2060 A>G và *IGFBP-7*. 4559 C>A (p.Leu189Met).

4. Các haplotype tạo nên từ 6 SNP tiềm năng (theo thứ tự trình bày ở kết luận 3) liên quan đến tính trạng tăng trưởng nhanh gồm: CO_H2 (ATGTGA), CO_H5 (TTCAGA), CO_H11 (TCCAAA), CO_H14 (TCGAAA); liên quan đến tính trạng tăng trưởng chậm gồm: CO_H1 (ATGTAA), CO_H3 (ACGTTA), CO_H7 (TTGTAA), CO_H8 (TTGTAC), CO_H9 (TTGTGA), CO_H10 (TTGAGA), CO_H13 (TCGTGA). Phân tích tổ hợp kiểu gen (diplotype) cho thấy 2 tổ hợp kiểu gen AT,CT,GC,AT,GA,AA (diplotype H2/H11) và AT,TT,GC,AT,GG,AA (diplotype H2/H5) liên quan đến tính trạng tăng trưởng nhanh.

Kiến nghị

Cần tiếp tục kiểm nghiệm sự liên quan giữa tính trạng tăng trưởng của cá tra nuôi và các chỉ thị SNP cũng như các haplotype, diplotype, các tổ hợp kiểu gen phát sinh trên một bộ mẫu lớn hơn nữa, hướng tới sử dụng chúng như các chỉ thị phân tử hỗ trợ chọn giống đối với tính trạng tăng trưởng của cá tra nuôi.

NHỮNG ĐÓNG GÓP MỚI CỦA LUẬN ÁN

- Đã phát hiện, phân tích, sàng lọc được 6 chỉ thị SNP tiềm năng trên 9 gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng của cá tra nuôi.

- Đã phân tích các haplotype, các tổ hợp kiểu gen liên quan đến các SNP có tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

1. Lê Thị Nguyên Bình, Nguyễn Thị Hoa, **Trần Thị Huyền Trang**, Nguyễn Thành Phương, Kim Thị Phương Oanh. *Phân tích cấu trúc gen mã hóa Insulin-like Growth Factor 2 (IGF2) ở cá tra nuôi (Pangasianodon hypophthalmus)*. Tạp chí Công nghệ sinh học, 2019, 17(3): 455-463.
2. **Trần Thị Huyền Trang**, Lê Thị Nguyên Bình, Nguyễn Thị Hoa, Trần Sơn Hoàng, Kim Thị Phương Oanh. *Phân tích cấu trúc một số gen Insulin-like Growth Factor Binding Protein (IGFBP) ở cá tra nuôi (Pangasianodon hypophthalmus)*. Tuyển tập Hội nghị Công nghệ Sinh học toàn quốc, 2019:17-22
3. **Trang Thi Huyen Tran**, Hoa Thi Nguyen, Binh Thi Nguyen Le, Phuc Huu Tran, Sang Van Nguyen, Oanh Thi Phuong Kim. *Characterization of single nucleotide polymorphism in IGF1 and IGF1R genes associated with growth traits in striped catfish (Pangasianodon hypophthalmus Sauvage, 1878)*. Aquaculture, 2021, 538: e736542.
4. **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim. *Significant association between a non-synonymous SNP in IGFBP5 gene and the growth of striped catfish (Pangasianodon hypophthalmus, Sauvage, 1878)*. Tạp chí Công nghệ sinh học, đã được chấp nhận đăng.
5. **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim. *Novel single nucleotide polymorphisms of Insulin-like Growth Factor Binding Protein 7 (IGFBP7) gene significantly associated with growth traits in striped catfish (Pangasianodon hypophthalmus Sauvage, 1878)*. Molecular Genetics and Genomics, accepted.