

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*)

LUẬN ÁN TIẾN SĨ SINH HỌC

Hà Nội – 2023

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*)**

LUẬN ÁN TIẾN SĨ SINH HỌC

**Chuyên ngành: Công nghệ sinh học
Mã số chuyên ngành: 9.42.02.01**

Xác nhận của Học viện
Khoa học và Công nghệ

**KT. GIÁM ĐỐC
PHÓ GIÁM ĐỐC**



Nguyễn Thị Trung

Hà Nội - 2023

Thầy hướng dẫn

A handwritten signature in blue ink, belonging to TS. Kim Thị Phương Oanh.

TS. Kim Thị Phương Oanh

LỜI CAM ĐOAN

Tôi xin cam đoan đây là công trình nghiên cứu của riêng tôi với sự hướng dẫn của TS. Kim Thị Phương Oanh. Những kết quả thu được của luận án là mới, trung thực và chưa từng được công bố trong bất kỳ công trình nào khác. Các kết quả công bố chung đã được cán bộ hướng dẫn và các đồng tác giả cho phép sử dụng trong Luận án.

Hà Nội, ngày 30 tháng 06 năm 2023

Tác giả luận án



Trần Thị Huyền Trang

LỜI CẢM ƠN

Trước tiên, tôi xin bày tỏ lòng biết ơn chân thành và sâu sắc nhất tới TS. Kim Thị Phương Oanh- Trưởng phòng Hệ gen học Môi trường, Viện Nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam. Tinh thần say mê, nghiêm túc trong nghiên cứu khoa học, sự hướng dẫn tận tình và sự động viên khích lệ của cô là động lực lớn lao giúp tôi không ngừng cố gắng, phấn đấu, vượt qua nhiều khó khăn để hoàn thành Luận án.

Tôi xin chân thành cảm ơn Ban Lãnh đạo Viện nghiên cứu hệ gen, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã luôn tạo mọi điều kiện thuận lợi trong suốt quá trình tôi học tập và thực hiện Luận án. Bên cạnh đó, tôi cũng xin gửi lời cảm ơn đến các nhà khoa học, các cán bộ nghiên cứu của Viện đã có những góp ý quý báu trong các buổi báo cáo, hội thảo, giúp tôi hoàn thiện Luận án của mình.

Tôi cũng xin chân thành cảm ơn ban Giám đốc, tập thể cán bộ Học viện Khoa học và Công nghệ, Viện Hàn lâm Khoa học và Công nghệ Việt Nam đã tạo điều kiện thuận lợi nhất cho tôi trong quá trình học tập và nghiên cứu tại đây.

Sau cùng, tôi xin gửi lời cảm ơn sâu sắc đến gia đình, người thân và bạn bè đã luôn tin tưởng và động viên tôi trong suốt quá trình học tập, nghiên cứu và hoàn thành Luận án.

Hà Nội, ngày 30 tháng 06 năm 2023

Tác giả luận án



Trần Thị Huyền Trang

MỤC LỤC

Trang

Trang phụ bìa	
LỜI CAM ĐOAN.....	i
LỜI CẢM ƠN	ii
MỤC LỤC	iii
DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT.....	vii
DANH MỤC CÁC BẢNG	ix
DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ	xi
MỞ ĐẦU	1
CHƯƠNG 1. TỔNG QUAN	4
1.1. Giới thiệu chung về cá tra nuôi <i>Pangasianodon hypophthalmus</i>	4
1.1.1. Vị trí phân loại và đặc điểm sinh học.....	4
1.1.2. Tình hình nuôi trồng và giá trị kinh tế	5
1.1.3. Các nghiên cứu liên quan đến chọn giống cá tra	5
1.2. Tăng trưởng (phát triển cơ) ở cá và các gen liên quan.....	8
1.2.1. Quá trình tăng trưởng ở cá.....	8
1.2.2. Điều hòa tăng trưởng ở cá.....	9
1.2.3. Một số gen liên quan đến tính trạng tăng trưởng đã được nghiên cứu ở cá xương.....	11
1.3. Vai trò, chức năng và đặc điểm cấu trúc của một số gen /protein thuộc hệ thống IGF.....	13
1.3.1. Vai trò, chức năng của một số protein thuộc hệ thống IGF đối với sự tăng trưởng của sinh vật	13

1.3.2. Đặc điểm cấu trúc một số gen và protein tương ứng thuộc hệ thống IGF của cá xương.....	16
1.4. Cơ sở tiền hành xác định các đa hình đơn nucleotide (SNP) trên một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng của cá tra nuôi	23
1.4.1. Vai trò quan trọng của SNP trong việc xác định các chỉ thị phân tử liên quan đến tính trạng quan tâm.....	23
1.4.2. Một số phương pháp xác định SNP (SNP genotyping) trên gen đích.....	24
1.4.3. Các nghiên cứu về SNP trên một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng ở cá xương	29
1.4.4. Các gen thuộc hệ thống IGF đã được xác định dựa trên thông tin hệ gen cá tra nuôi.....	31
CHƯƠNG 2. NGUYÊN VẬT LIỆU VÀ PHƯƠNG PHÁP.....	34
2.1. Nguyên vật liệu	34
2.2. Phương pháp	36
2.2.1. Phân tích cấu trúc của các gen đích	36
2.2.2. Tách chiết DNA tổng số.....	37
2.2.3. Khuếch đại các gen bằng phản ứng PCR.....	37
2.2.4. Giải trình tự các gen IGF1, IGF2, IGF1R, IGFBP-1, -2, -3, -5, -6, -7 bằng phương pháp Sanger	47
2.2.5. Kiểm tra tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham chiếu.....	48
2.2.6. Phát hiện và sàng lọc các SNP trên các gen đích trên bộ mẫu khởi tạo	48
2.2.7. Xác định các SNP được sàng lọc trên bộ mẫu kiểm nghiệm bằng phương pháp kéo dài một nucleotide (SBE).....	50
2.2.8. Phân tích dữ liệu	53
CHƯƠNG 3. KẾT QUẢ.....	55

3.1. Phân tích cấu trúc của một số gen/protein thuộc hệ thống IGF của cá tra nuôi .	55
3.1.1. Các gen mã hóa IGF (<i>IGF1</i> và <i>IGF2</i>) và các protein tương ứng	55
3.1.2. Gen mã hóa thụ thể <i>IGF1</i> (<i>IGF1R</i>) và protein tương ứng	58
3.1.3. Một số gen mã hóa <i>IGFBP</i> và các protein tương ứng	60
3.2. Xác định tính xác thực của trình tự giải mã bằng Sanger so với trình tự tham chiếu	63
3.3. Phát hiện, sàng lọc SNP trên các gen đích thuộc hệ thống IGF trên bộ mẫu khởi tạo	64
3.3.1. Phát hiện và sàng lọc SNP trên gen <i>IGF1</i>	64
3.3.2. Phát hiện và sàng lọc SNP trên gen <i>IGF2</i>	66
3.3.3. Phát hiện và sàng lọc SNP trên gen <i>IGF1R</i>	68
3.3.4. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-1</i>	70
3.3.5. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-2</i>	71
3.3.6. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-3</i>	72
3.3.7. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-5</i>	74
3.3.8. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-6</i>	75
3.3.9. Phát hiện và sàng lọc SNP trên gen <i>IGFBP-7</i>	78
3.4. Phân tích sự liên quan đến tính trạng tăng trưởng của các SNP được sàng lọc .	82
3.4.1. SNP genotyping trên các cá thể thuộc bộ mẫu kiểm nghiệm	82
3.4.2. Phân tích mối liên quan giữa SNP được sàng lọc và tính trạng tăng trưởng của cá tra nuôi	84
CHƯƠNG 4. THẢO LUẬN	98
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.....	98
4.2. Sự đa dạng di truyền của các gen thuộc hệ thống IGF của cá tra nuôi	100

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	105
4.4. Phát triển chỉ thị phân tử SNP nhằm ứng dụng chọn giống cá tra theo hướng tăng trưởng.....	108
KẾT LUẬN	110
KIẾN NGHỊ.....	111
DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ.....	112
TÀI LIỆU THAM KHẢO	113
PHỤ LỤC.....	127

DANH MỤC CÁC KÝ HIỆU, CÁC CHỮ VIẾT TẮT

Tên viết tắt	Tên đầy đủ	Nghĩa tiếng Việt dùng trong luận án
AFLP	Amplified fragment length polymorphism	Đa hình chiều dài các đoạn khuếch đại
ALS	Acid labile subunit	Tiểu đơn vị axit không bền
Alt	Alternative	Thay thế
ATP	Adenosine triphosphate	
BLAST	Basic Local Alignment Search Tool	Công cụ tìm kiếm so sánh trình tự cơ bản
CDS	Coding sequence	Vùng mã hóa
CpG		Đảo CpG
Cys	Cysteine	
EBV	Estimated breeding value	Giá trị chọn giống ước đoán
FRET	Fluorescence resonance energy transfer	Chuyển đổi năng lượng cộng hưởng huỳnh quang
GF	Growth factor	Nhân tố tăng trưởng
GH	Growth hormone	Hormone tăng trưởng
GHIH	Growth hormone-inhibiting hormone	Hormone ức chế hormone tăng trưởng
GHRH	Growth hormone-releasing hormone	Hormone giải phóng hormone tăng trưởng
H	Haplotype	Kiểu gen đơn bội
IGF	Insulin like growth factor	Nhân tố tăng trưởng tương tự insulin
IGFBP	Insulin like growth factor binding protein	Protein bám với IGF
IGFBP-rP	IGFBP-related protein	Protein liên quan đến IGFBP
IGFR	Insulin like growth factor receptor	Thụ thể IGF
LD	Linkage disequilibrium	Mất cân bằng liên kết
LOH	Loss of heterozygosity	Mất tính dị hợp tử
MAF	Minor allele frequency	Tần số alen thiểu số

MAS	Marker-assisted selection	Chọn giống nhờ chỉ thị
MRFs	Myogenic regulatory factors	Các nhân tố điều hòa cơ
NCBI	National Center for Biotechnology Information	Trung tâm Thông tin Công nghệ sinh học Quốc gia
NGS	Next generation sequencing	Giải trình tự gen thế hệ mới
NN	Non-identified	Không xác định được
NST	Chromosome	Nhiễm sắc thể
PIC	Polymorphism information content	Thông tin đa hình
QTL	Quantitative trait loci	Locus quy định tính trạng số lượng
RAD	Restriction site associated DNA	Vị trí cắt của enzym giới hạn trên DNA
RAPD	Random amplified polymorphism DNA	Nhân ngẫu nhiên DNA đa hình
Ref	Reference	Tham chiếu
RNA-seq	RNA sequencing	Giải trình tự RNA
SAP	Shrimp alkaline phosphatase	
SBE	Single base extension	Kéo dài một nucleotide
SNP	Single nucleotide polymorphism	Đa hình đơn nucleotide
SP	Signal peptide	Peptit tín hiệu
MAP	Simple sequence repeat	Đoạn lặp trình tự đơn giản
TTC		Tăng trưởng chậm
TTN		Tăng trưởng nhanh
UTR	Untranslated region	Vùng không dịch mã
WGS	Whole genome sequencing	Giải trình tự toàn bộ hệ gen

DANH MỤC CÁC BẢNG

<i>Bảng 1.1.</i> Các gen thuộc hệ thống IGF của cá tra nuôi.....	32
<i>Bảng 1.2.</i> Số định danh trình tự gen, mRNA, protein trên NCBI của một số gen trong hệ thống IGF của cá tra nuôi được lựa chọn để phân tích cấu trúc	33
<i>Bảng 2.1.</i> Thông tin 20 cá thể cá tra nuôi trong bộ mẫu khởi tạo	35
<i>Bảng 2.2.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGF1</i>	38
<i>Bảng 2.3.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGF2</i>	39
<i>Bảng 2.4.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGF1R</i>	41
<i>Bảng 2.5.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-1</i>	42
<i>Bảng 2.6.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-2</i>	43
<i>Bảng 2.7.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-3</i>	45
<i>Bảng 2.8.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-5</i>	45
<i>Bảng 2.9.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-6</i>	46
<i>Bảng 2.10.</i> Danh sách các cặp môi được thiết kế nhằm nhân các đoạn gen <i>IGFBP-7</i> ..	47
<i>Bảng 2.11.</i> Trình tự môi nhân các đoạn chứa SNP được sàng lọc và môi SBE cho phản ứng kéo dài một nucleotide (SBE)	51
<i>Bảng 3.1.</i> Danh sách các SNP được phát hiện trên gen <i>IGF1</i>	66
<i>Bảng 3.2.</i> Danh sách các SNP được phát hiện trên gen <i>IGF2</i>	67
<i>Bảng 3.3.</i> Danh sách các SNP được phát hiện trên gen <i>IGF1R</i>	69
<i>Bảng 3.4.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-1</i>	70
<i>Bảng 3.5.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-2</i>	71
<i>Bảng 3.6.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-3</i>	73
<i>Bảng 3.7.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-5</i>	75
<i>Bảng 3.8.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-6</i>	77
<i>Bảng 3.9.</i> Danh sách các SNP được phát hiện trên gen <i>IGFBP-7</i>	80
<i>Bảng 3.10.</i> Danh sách các SNP được sàng lọc trên bộ mẫu khởi tạo	82
<i>Bảng 3.11.</i> Phân tích mối liên quan giữa SNP 13680 A>T trên gen <i>IGF1</i> với tính trạng tăng trưởng của cá tra nuôi	85

<i>Bảng 3.12.</i> Phân tích mối liên quan giữa 3 SNP trên gen <i>IGF1R</i> với tính trạng tăng trưởng của cá tra nuôi.	86
<i>Bảng 3.13.</i> Phân tích sự mất cân bằng liên kết (linkage disequilibrium (LD)) của 3 SNP trên gen <i>IGF1R</i> thông qua giá trị D' và R^2	87
<i>Bảng 3.14.</i> Phân tích mối liên quan giữa haplotype phát sinh từ 3 SNP của gen <i>IGF1R</i> lên tính trạng tăng trưởng của cá tra nuôi.	88
<i>Bảng 3.15.</i> Phân tích mối liên quan giữa SNP 704 C>G (p.Leu8Val) trên gen <i>IGFBP-3</i> với tính trạng tăng trưởng của cá tra nuôi	89
<i>Bảng 3.16.</i> Phân tích mối liên quan giữa SNP 525 T>A (p.Val16Glu) trên gen <i>IGFBP-5</i> với tính trạng tăng trưởng của cá tra nuôi.	90
<i>Bảng 3.17.</i> Phân tích mối liên quan giữa SNP 2278 C>A trên gen <i>IGFBP-6</i> với tính trạng tăng trưởng của cá tra nuôi	91
<i>Bảng 3.18.</i> Phân tích mối liên quan giữa 3 SNP trên gen <i>IGFBP-7</i> với tính trạng tăng trưởng của cá tra nuôi.	92
<i>Bảng 3.19.</i> Phân tích sự mất cân bằng liên kết (linkage disequilibrium (LD)) của 3 SNP trên gen <i>IGFBP-7</i> thông qua giá trị D' và R^2	93
<i>Bảng 3.20.</i> Phân tích mối liên quan giữa haplotype phát sinh từ 3 SNP của gen <i>IGFBP-7</i> lên tính trạng tăng trưởng	94
<i>Bảng 3.21.</i> 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	95
<i>Bảng 3.22.</i> 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	97

DANH MỤC CÁC HÌNH VẼ, ĐỒ THỊ

<i>Hình 1.1.</i> Cá tra nuôi <i>Pangasianodon hypophthalmus</i>	4
<i>Hình 1.2.</i> Sự phát triển cơ sau giai đoạn phôi ở cá xương	9
<i>Hình 1.3.</i> Mô hình điều hòa phát triển cơ ở cá	10
<i>Hình 1.4.</i> Hệ thống IGF ở động vật có vú	13
<i>Hình 1.5.</i> Cấu trúc tiền thân IGF và cấu trúc bậc ba của protein IGF ở cá xương.	17
<i>Hình 1.6.</i> Cấu trúc tetramer IGF1R gồm 2 chuỗi dị dimer alphabeta.....	19
<i>Hình 1.7.</i> Cấu trúc gen, protein IGF1Ra (A) và IGF1Rb (B) của các loài cá	21
<i>Hình 1.8.</i> Cấu trúc gen và protein của các IGFBP	22
<i>Hình 1.9.</i> Nguyên lý kỹ thuật PCR đặc hiệu alen dựa trên FRET	26
<i>Hình 1.10.</i> Hình ảnh minh họa quá trình thực hiện thí nghiệm SBE sử dụng SNaPShot	27
<i>Hình 2.1.</i> Nội dung nghiên cứu và các nhóm phương pháp nghiên cứu tương ứng.....	36
<i>Hình 2.2.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGF1</i>	38
<i>Hình 2.3.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGF2</i>	39
<i>Hình 2.4.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGF1R</i>	40
<i>Hình 2.5.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGFBP-1</i>	42
<i>Hình 2.6.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGFBP-2</i>	43
<i>Hình 2.7.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGFBP-3</i> , <i>IGFBP-5</i> và <i>IGFBP-6</i>	44
<i>Hình 2.8.</i> Vị trí các cặp môi được thiết kế trên trình tự gen <i>IGFBP-7</i>	46
<i>Hình 2.9.</i> Tiêu chí sàng lọc các SNP được phát hiện ở bộ mẫu khởi tạo.....	49
<i>Hình 3.1.</i> Kết quả phân tích cấu trúc gen và protein IGF1 của cá tra nuôi.....	55
<i>Hình 3.2.</i> Kết quả phân tích cấu trúc gen và protein IGF2 của cá tra nuôi.....	57
<i>Hình 3.3.</i> Kết quả phân tích cấu trúc gen và protein IGF1R của cá tra nuôi	59
<i>Hình 3.4.</i> Kết quả phân tích cấu trúc gen, protein IGFBP-1 và IGFBP-2	60
<i>Hình 3.5.</i> Kết quả phân tích cấu trúc gen, protein IGFBP-3 (A), IGFBP-5 (B) và IGFBP-6 (C).....	61
<i>Hình 3.6.</i> Kết quả phân tích cấu trúc gen và protein IGFBP-7.....	63

<i>Hình 3.7.</i> Kết quả phát hiện <i>IGF1.13680 A>T</i> trên gen <i>IGF1</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu. ..65	65
<i>Hình 3.8.</i> Kết quả phát hiện <i>IGF1R.83894 A>G</i> trên gen <i>IGF1R</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu. ..68	68
<i>Hình 3.9.</i> Kết quả phát hiện <i>IGFBP-3.704 C>G</i> trên gen <i>IGFBP-3</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu.72	72
<i>Hình 3.10.</i> Kết quả phát hiện <i>IGFBP-5.525 T>A</i> trên gen <i>IGFBP-5</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu.74	74
<i>Hình 3.11.</i> Kết quả phát hiện <i>IGFBP-6. 2278 C>A</i> trên gen <i>IGFBP-6</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu.....76	76
<i>Hình 3.12.</i> Kết quả phát hiện <i>IGFBP-7.4559 C>A</i> trên gen <i>IGFBP-7</i> thông qua so sánh các trình tự tương ứng được giải mã bằng phương pháp Sanger với trình tự tham chiếu79	79
<i>Hình 3.13.</i> Kết quả điện di mao quản sản phẩm SNaPshot trên một mẫu đại diện của nhóm môi SBE thứ nhất sau khi được phân tích trên phần mềm GeneMapper.83	83
<i>Hình 3.14.</i> Kết quả điện di mao quản SNaPshot trên một mẫu đại diện của nhóm môi SBE thứ hai sau khi được phân tích trên phần mềm GeneMapper.....84	84
<i>Hình 4.1.</i> Kết quả phân tích vị trí của các SNP được phát hiện trên các gen.....100	100
<i>Hình 4.2.</i> Tỷ lệ các SNP được sàng lọc trên tổng số SNP được phát hiện trên các gen.....102	102
<i>Hình 4.3.</i> Kết quả so sánh các trình tự protein <i>IGFBP-7</i> của cá tra nuôi (<i>Pangasianodon hypophthalmus</i>) và các loài cá xương khác.....103	103

MỞ ĐẦU

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, với sản lượng hàng năm chiếm khoảng 30% tổng sản lượng thủy sản cả nước. Ngay trong giai đoạn dịch bệnh Covid, Việt Nam cũng định hướng tới việc xuất khẩu kim ngạch cá tra hơn 1,6 tỷ USD, cho thấy rõ vai trò của cá tra trong xuất khẩu thủy sản nói riêng và kinh tế nói chung. 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 ảnh hưởng không nhỏ đến các quần thể cá tra. Để phát triển bền vững ngành sản xuất cá tra, bên cạnh việc cải thiện kỹ thuật canh tác, nguồn thức ăn ổn định, một trong những nhiệm vụ cấp thiết, quan trọng hơn cả cần thực hiện đầu tiên chính là nâng cao chất lượng nguồn giống.

Tăng trưởng là tính trạng được quan tâm hàng đầu trong các chương trình chọn giống thủy sản. Chương trình chọn giống cá tra theo hướng tăng trưởng nhanh dựa trên di truyền số lượng đã được thực hiện từ năm 2001. Việc kết hợp giữa di truyền số lượng và di truyền phân tử đã giúp rút ngắn thời gian chọn giống, đặt nền móng ban đầu cho công tác chọn giống và nghiên cứu đa dạng sinh học của giống thủy sản có giá trị kinh tế cao này. Trong chọn giống nhờ chỉ thị (Marker - assisted selection (MAS)), trên thế giới, chỉ thị đa hình đơn nucleotide (Single-nucleotide polymorphism (SNP)) đã được ứng dụng thành công ở các loài cá, giáp xác và hai mảnh vỏ. Với thành tựu giải mã hệ gen cá tra nuôi được công bố gần đây, việc nghiên cứu sự đa dạng di truyền của các gen liên quan đến tính trạng tăng trưởng của cá tra nuôi hứa hẹn mang đến triển vọng khai thác nhanh thông tin hệ gen, hướng tới ứng dụng sớm nhất có thể các chỉ thị SNP trong công tác chọn giống cá tăng trưởng nhanh.

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 dẫn tới quá trình tăng trưởng, biệt hóa và tăng sinh tế bào. Mặc dù việc tìm kiếm các

chỉ thị SNP trên các gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng đã được tiến hành ở một số loài cá xương, tuy nhiên số lượng các nghiên cứu vẫn còn hạn chế và chưa từng được triển khai trên đối tượng cá tra nuôi.

Dựa vào những cơ sở nêu trên, Luận án quyết định nghiên cứu đa hình SNP trên một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng ở cá tra nuôi, 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.

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.

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.

Nội dung nghiên cứu 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.

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 phát sinh từ các SNP có tiềm năng liên quan đến các SNP có tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

Đề tài có ý nghĩa khoa học và thực tiễn cao khi cung cấp dữ liệu về đa hình SNP của 9 gen thuộc hệ thống IGF, chỉ ra 6 chỉ thị SNP tiềm năng và các haplotype, tổ hợp kiểu gen liên quan đến tính trạng tăng trưởng của cá tra nuôi, hướng tới ứng dụng trong chương trình chọn giống. Những kết quả thu được của luận án đã đóng góp vào quá trình phát triển chương trình chọn giống cá tra nuôi theo hướng tăng trưởng nhanh.

CHƯƠNG 1. TỔNG QUAN

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

1.1.1. Vị trí phân loại và đặc điểm sinh học

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). Ở Thái Lan, cá tra còn được gặp ở lưu vực sông Chao Phraya [1].



Hình 1.1. Cá tra nuôi *Pangasianodon hypophthalmus*

Cá tra nuôi *P. hypophthalmus* là cá da trơn thân dài, sau dẹt và không có vảy; đầu tương đối nhỏ, mắt lớn, miệng rộng, răng nhỏ, sắc nhọn; vây màu xám đen hoặc đen, tia vây lưng 6, màu sắc cơ thể đồng nhất với màu xám nhưng đôi khi có sắc xanh lục và các mặt bên màu bạc, con non có một sọc đen dọc theo đường bên. Cá có khả năng di cư mạnh mẽ với quãng đường vài trăm km từ vùng thượng nguồn đến nơi sinh sản, nơi kiếm ăn và ương dưỡng ở hạ nguồn.

Cá trưởng thành có thể đạt chiều dài cơ thể tới 130 cm và khối lượng lên đến 44 kg. Cá tra nuôi thể hiện tính sống đáy với phạm vi pH từ 6,5 đến 7,5 và nhiệt độ từ 22 đến 26°C và thể hiện tính ăn tạp, với thức ăn gồm tảo, thực vật bậc cao, động vật phù du và côn trùng, cá lớn còn ăn trái cây, động vật giáp xác và cá.

Về sinh sản, trong điều kiện nuôi nhốt cũng như trong tự nhiên, con cái mất khoảng ba năm còn con đực mất khoảng hai năm để thành thực khi đạt cân nặng khoảng 3 kg. Mùa vụ thành thực của cá trong tự nhiên bắt đầu từ tháng 5-6 dương lịch. Cá có tập tính di cư để tự nhiên trên những khúc sông có điều kiện sinh thái phù hợp, sau khi trứng được đẻ ra khoảng 24 giờ sẽ nở thành cá bột và trôi về hạ nguồn. Một con cái trưởng thành nặng 10 kg có thể đẻ hơn một triệu quả trứng. Cá tra tự

nhiên thường sinh sản hai lần mỗi năm, nhưng trong điều kiện nuôi lồng ở Việt Nam đã ghi nhận được lần sinh sản thứ hai của cá tra chỉ cách từ 6 đến 17 tuần sau lần sinh sản đầu tiên [2].

1.1.2. Tình hình nuôi trồng và giá trị kinh tế

Do có thịt trắng, không mùi, bổ dưỡng, chứa nhiều các thành phần vitamin A, D, E, các axit béo không no thiết yếu cho cơ thể, không chứa cholesterol, hương vị sau khi nấu thơm ngon, giá cả phải chăng nên cá tra được ưa chuộng sử dụng, nhất là ở châu Âu và châu Mỹ. Căn cứ Quyết định số 50/2018/QĐ-TTg ngày 13 tháng 12 năm 2018 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. Theo báo cáo từ Hiệp hội chế biến và xuất khẩu thủy sản Việt Nam (VASEP), năm 2021 dù đối mặt với dịch Covid-19, xuất khẩu cá tra vẫn tăng 8,4% so với 2020, kim ngạch xuất khẩu đạt trên 1,61 tỷ USD [3]. Theo Bộ Nông nghiệp và Phát triển Nông thôn (Bộ NN&PTNT), trong năm 2022, ngành cá tra dự kiến kế hoạch sản xuất với sản lượng cá thương phẩm đạt 1,6-1,7 triệu tấn; kim ngạch xuất khẩu đạt trên 1,6 tỷ USD [4]. Cụ thể, kế hoạch năm 2022 của Đồng Tháp, một trong ba vùng nuôi cá tra lớn nhất đồng bằng sông Cửu Long, là tăng diện tích thả nuôi nhằm tăng sản lượng lên 495.000 tấn (tăng 1,8% so với năm 2021), ngoài ra, tăng sản lượng cá tra bột 24.000 triệu con (tăng 28,3% so với năm 2021) và sản lượng cá tra giống 1.750 triệu con (tăng 55,4% so với năm 2021) [4]. Tính chung quý I năm 2022, sản lượng cá tra đạt 342,6 nghìn tấn, tăng 6,5% so với cùng kỳ năm trước. Tính đến tháng 5/2022, xuất khẩu cá tra cả nước ước đạt hơn 1,1 tỷ USD, tăng 97% so với cùng kỳ năm 2021, các doanh nghiệp chế biến và xuất khẩu cá tra đang tích cực mở rộng thị trường mới song song với xuất khẩu sang các thị trường truyền thống như Mỹ, châu Âu [5]. Để đạt được mục tiêu đề ra, Bộ NN&PTNT 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 [4]. Một trong số những nhiệm vụ cấp thiết chính là triển khai và ứng dụng các kết quả thu được từ các chương trình chọn giống hướng tới nâng cao tính trạng tăng trưởng cũng như chống chịu bệnh trên cá tra nuôi.

1.1.3. Các nghiên cứu liên quan đến chọn giống cá tra

Trong những năm gần đây, ở nước ta, các nghiên cứu nhằm nâng cao chất lượng di truyền và kiểm soát dịch bệnh ở một số giống thủy sản đã ngày càng được

chú ý đến. Trong đó, các nghiên cứu đã và đang tiến hành trên cá tra đã đặt nền móng ban đầu cho công tác chọn giống và nghiên cứu đa dạng sinh học của giống thủy sản có giá trị kinh tế cao này.

Các tác giả Phạm Anh Tuấn và Nguyễn Hữu Ninh (2003) đã sử dụng 4 microsatellite tìm kiếm mối tương quan giữa marker phân tử với màu sắc thịt của cá tra [6]. Năm 2015, tác giả Nguyễn Minh Thành và cộng sự đã nghiên cứu hệ gen chức năng ở mô thận của cá tra nuôi nhằm xác định các gen tiềm năng liên quan đến khả năng chịu mặn của cá tra [7]. Một số nghiên cứu đa dạng di truyền sử dụng các marker phân tử như RAPD và AFLP [8], [9] và chỉ thị phân tử microsatellite đánh giá đa dạng di truyền quần đàn [10] đã được tiến hành trên cá tra.

Các chương trình chọn tạo giống cá tra nuôi được Viện Nghiên cứu Nuôi trồng Thủy sản 2 bắt đầu từ năm 2001. Từ đó cho đến năm 2016, các đề tài liên quan đến các chương trình chọn tạo giống được tiến hành xuyên suốt và có tính kế thừa, bao gồm: 1) chương trình chọn giống đầu tiên (2001 – 2005) trên cá tra theo tính trạng tăng trưởng bằng phương pháp chọn lọc cá thể, với sự hỗ trợ kinh phí của Hợp phần Phát triển Nuôi trồng Thủy sản (SUFA), 2) đề tài nghiên cứu cấp Bộ ‘Chọn giống cá tra nhằm tăng tỉ lệ philê bằng chọn lọc gia đình’ (2006 – 2008) với hai tính trạng chọn lọc là tăng trưởng và tỉ lệ philê, 3) đề tài cấp nhà nước thuộc Chương trình Công nghệ Sinh học Nông nghiệp Thủy sản ‘Đánh giá hiệu quả chọn giống cá tra về tăng trưởng và tỉ lệ philê’ tiếp tục được thực hiện trong giai đoạn 2010 – 2012, 4) đề tài ‘Ứng dụng di truyền phân tử và di truyền số lượng chọn giống nâng cao tốc độ sinh trưởng cá tra’ cũng thuộc chương trình Công nghệ Sinh học Nông nghiệp Thủy sản được thực hiện trong giai đoạn 2013 – 2016 tập trung nghiên cứu vào tính trạng tăng trưởng. Cụ thể, các đề tài áp dụng lý thuyết di truyền số lượng và một phần di truyền phân tử vào chọn giống cá tra nâng cao tốc độ tăng trưởng qua 3 thế hệ. Quần thể chọn giống ban đầu (G0) có biến di truyền cao được thành lập dựa trên cá bố mẹ có nguồn gốc từ 3-4 trại giống khác nhau và được đánh bắt từ tự nhiên trước đó. Phương pháp ghép phối thứ bậc và giai thừa một phần được áp dụng thế hệ G1-G3 cho phép ước tính các thông số di truyền chính xác và từ đó áp dụng cho chọn lọc chính xác. Phương pháp đánh dấu từ PIT (Passive Integrated Transponder) phân biệt đến từng cá thể được áp dụng nhằm phân biệt đến từng cá thể để so sánh tốc độ tăng trưởng của cá thể và phối tránh cận huyết trong các đề tài. Các phần mềm thống kê và các thuật toán

mới và hiện đại luôn được cập nhật và áp dụng như Excel, SAS, R và ASReml được dùng để quản lý và xử lý số liệu, để ước tính các thông số di truyền chính xác và từ đó áp dụng cho chọn lọc chính xác hơn. Đặc biệt, phương pháp đóng góp tối ưu OC (Optimum Contribution) được áp dụng nhằm xác định các cá thể và gia đình cần chọn lọc và phối sản xuất gia đình tạo thế hệ tiếp theo để tối ưu hiệu quả chọn lọc— tăng trưởng nhanh và giới hạn hệ số cận huyết ở mức giới hạn được áp dụng từ thế hệ G2 trở đi. Áp dụng phương pháp chọn lọc kết hợp (combine selection), tức chọn lọc dựa trên biến dị di truyền giữa (between family) và trong nội bộ gia đình (within family) nhằm tăng độ chính xác và hiệu quả chọn lọc. Hiệu quả chọn lọc tính trạng tăng trưởng thông qua khối lượng lúc thu hoạch được so sánh giữa nhóm đã qua chọn lọc so với nhóm đối chứng cùng thế hệ chọn giống và nhóm cá có bố mẹ từ tự nhiên (chưa qua chọn lọc). Kết quả thu được hệ số di truyền ước tính cho các tính trạng tăng trưởng ở thế hệ thứ 2 (G2) nằm ở mức trung bình (0,24 – 0,30) [11], [12], tính trạng khối lượng thu hoạch được chọn lọc đến thế hệ thứ 3 với hệ số di truyền ước tính ở mức trung bình-cao (0,34-0,52) và hệ số di truyền thực tế ở mức trung bình (0,24-0,38) [13]. Chương trình chọn tạo giống cũng đưa ra danh sách chọn lọc và phối tránh cận huyết và chọn tạo được quần thể chọn giống nâng cao tốc độ tăng trưởng thế hệ G3 (G3-cộng gộp) với hệ số di truyền ước tính tính trạng khối lượng thu hoạch cao (0,51), đảm bảo hiệu quả chọn lọc cao nhất nhưng vẫn hạn chế tỷ lệ cận huyết ở mức xác định trước. Từ đàn G3-cộng gộp, 1.230 cá thể G3 thuộc 157 gia đình được chọn (G3-chọn lọc) cho chọn giống tiếp theo làm cá bố mẹ để sản xuất quần thể cá tra tăng trưởng cao tiếp theo và có hiệu quả chọn lọc thực tế cao hơn 20,4% so với đàn cá chưa qua chọn lọc [14], [15]. Tiếp đó, đề tài "Nghiên cứu chọn giống cá tra (*Pangasianodon hypophthalmus*) nâng cao sinh trưởng" thực hiện trong giai đoạn 2019-2022 đã chọn lọc thêm một thế hệ nữa (G4-tăng trưởng) nhằm nâng cao hơn nữa hiệu quả chọn lọc của tính trạng tăng trưởng.

Tác giả Kim Thị Phương Oanh và nhóm nghiên cứu (2018) đã giải mã toàn bộ hệ gen (genome) của một cá thể cá tra nuôi bằng kỹ thuật giải trình tự thế hệ mới (hệ thống Illumina), dữ liệu thu được được sử dụng để lắp ráp bộ gen với kích thước hệ gen nhân ước tính khoảng 700 Mbp. Kết hợp với dữ liệu transcriptome từ các mô cơ quan và các giai đoạn phát triển khác nhau, genome cá tra đã được dự đoán mô hình gen và chú giải, dự đoán có 28600 gen mã hóa protein. Đây là dữ liệu genome đầu

tiên của cá tra nuôi, có thể dùng làm trình tự tham chiếu cho các nghiên cứu sâu sau này [16]. Nhóm nghiên cứu cũng sử dụng kỹ thuật RNA-seq để giải mã transcriptome từ mô cơ và mô não của của hai nhóm cá tăng trưởng nhanh và tăng trưởng chậm, sau đó phân tích, so sánh, tìm kiếm và kiểm nghiệm được các SNP trên hệ gen biểu hiện liên quan đến tính trạng tăng trưởng ở cá tra nuôi [17], [18].

Tác giả Marnis và cộng sự (2018) cũng kiểm nghiệm sự liên quan của 5 chỉ thị microsatellite với các tính trạng tăng trưởng của cá tra nuôi, kết quả cho thấy một chỉ thị liên quan tới chiều dài cơ thể, một chỉ thị liên quan đến khối lượng cơ thể và một chỉ thị liên quan đến cả hai tính trạng trên. Các kiểu gen tạo nên từ 3 chỉ thị trên cũng cho thấy mối liên quan đến các tính trạng tăng trưởng của cá tra nuôi, có thể sử dụng làm các chỉ thị phân tử hỗ trợ chọn giống theo hướng tăng trưởng [19].

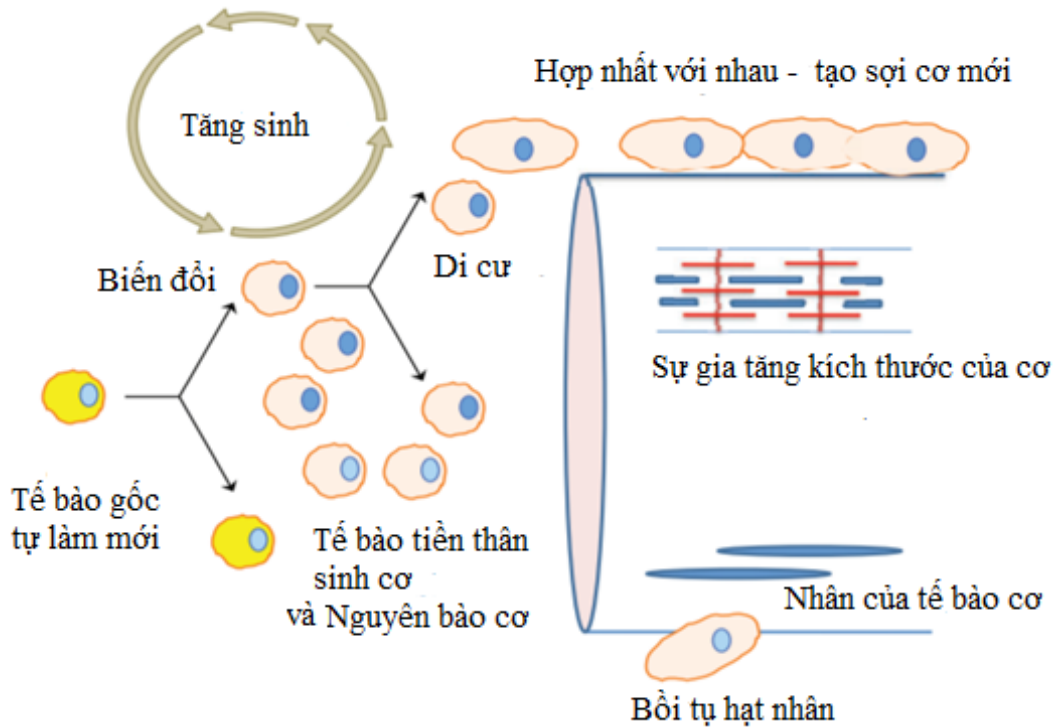
1.2. Tăng trưởng (phát triển cơ) ở cá và các gen liên quan

1.2.1. Quá trình tăng trưởng ở cá

Phát triển cơ là một trong những quá trình tăng trưởng quan trọng ở cá, bắt đầu từ giai đoạn sớm của phôi với sự điều khiển của các phân tử có trong lòng đỏ trứng và các yếu tố kích thích khác từ môi trường như nhiệt độ và oxy. Nhiều quần thể các tế bào tiền thân sinh cơ phân bố trong các đốt phôi và lớp tế bào bên ngoài đã tham gia vào quá trình tạo thành các sợi cơ ở các giai đoạn phôi, cá bột, cá con và cá trưởng thành. Nhiều protein tín hiệu và các nhân tố phiên mã cần thiết cho quá trình này đã được nghiên cứu. Trong tất cả các trường hợp, sự phát triển cơ đều bao gồm sự tăng sinh của nguyên bào cơ, di cư, hợp nhất và biệt hóa cuối cùng. Ngoài ra, sự phát triển của hệ tuần hoàn và hệ bạch huyết cùng các cơ quan tiêu hóa, nội tiết, thần kinh và miễn dịch làm tăng cường trao đổi thông tin giữa các mô và với môi trường bên ngoài khiến cho quá trình điều hòa tăng trưởng cơ trở nên phức tạp hơn. Cá xương thường thể hiện một mô hình tăng trưởng không xác định, với kích thước cơ thể và khối lượng cơ bấp bênh cho đến khi sự tử vong hoặc lão hóa xảy ra. Sự gia tăng đáng kể khối lượng cơ giữa giai đoạn phôi và giai đoạn trưởng thành đòi hỏi phải sản xuất sợi cơ liên tục cho đến khi đạt được 40 đến 50% chiều dài cơ thể tối đa [20].

Sự phát triển cơ sau giai đoạn phôi ở cá xương được minh họa trong Hình 1.2. Các tế bào gốc có thể tự làm mới (stem cell renewal) hoặc biến đổi thành các tế bào tiền thân sinh cơ (myogenic precursor cells (MPCs)) (determination). Các MPCs có thể tăng sinh (proliferation) và di cư (migration) xuyên qua các cơ trước khi hoàn

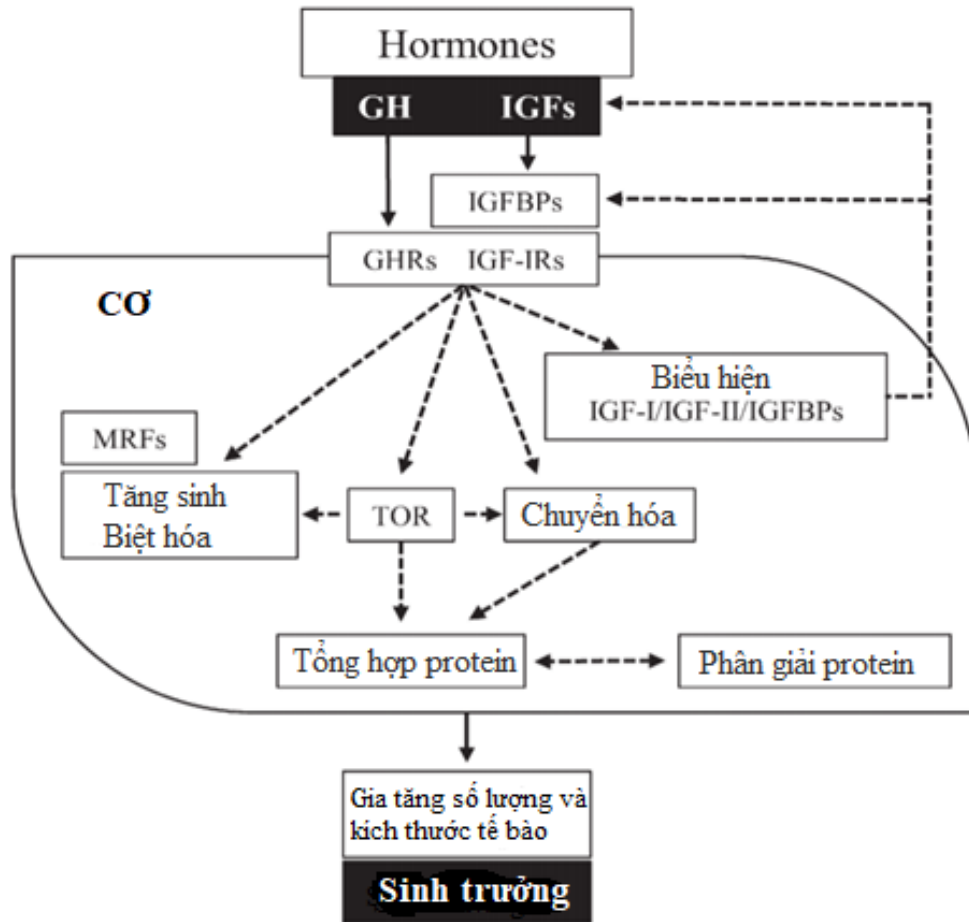
thành chu trình tế bào và bước vào giai đoạn biệt hóa cuối cùng. Các tế bào tiền thân sinh cơ hoặc các nguyên bào cơ (committed myoblast) sẽ biến đổi theo một trong hai hướng: hoặc hợp nhất với nhau (fusion) để tạo thành các sợi cơ ngắn trên bề mặt của các sợi cơ, hoặc được hấp thụ vào trong các sợi cơ giúp chúng dài ra và tăng cường kích thước trong quá trình tăng trưởng (bồi tụ hạt nhân - nuclear accretion).



Hình 1.2. Sự phát triển cơ sau giai đoạn phôi ở cá xương [16]

1.2.2. Điều hòa tăng trưởng ở cá

Hệ trục hormone liên quan đến tăng trưởng gồm: hormone tăng trưởng (Growth hormone (GH)), hormone giải phóng hormone tăng trưởng (Growth hormone-releasing hormone (GHRH)), hormone ức chế hormone tăng trưởng (Growth hormone-inhibiting hormone (GHIH)), các nhân tố tăng trưởng tương tự insulin (Insulin-like growth factors) (IGF1 và IGF2), một số protein mang và thụ thể khác đóng vai trò quan trọng trong việc điều hòa tăng trưởng, đặc biệt là quá trình phát triển cơ ở cá xương. Trong đó, GH điều hòa sự tổng hợp cơ, xác định sự tăng trưởng cơ ở cả hai quá trình: hyperplasia (tăng sản- tăng kích thước của cơ bắp nhờ sự gia tăng về số lượng tế bào) và hypertrophy (gia tăng kích thước của tế bào sợi cơ). Các nhân tố tăng trưởng tương tự insulin, đặc biệt là IGF1 đóng vai trò quan trọng trong điều hòa sự tạo cơ, bao gồm sự hoạt hóa, tăng sinh và biệt hóa [21].



Hình 1.3. Mô hình điều hòa phát triển cơ ở cá [17]

Cơ chế điều hòa sự phát triển cơ của hệ trục này như sau: các tín hiệu từ môi trường bên ngoài sau khi được thu nhận từ các tuyến (ví dụ tuyến tùng ở trên não bộ) sẽ được xử lý ở vùng dưới đồi, dẫn tới điều hòa các hormone như GHRH và GHIH. GHRH và GHIH được tiết ra tĩnh mạch của ở thùy trước tuyến yên và bám vào các thụ thể đặc hiệu kích thích hoặc ức chế sự tiết hormone GH. GH bám protein đặc hiệu và được tiết vào trong máu, rồi tiếp tục bám các thụ thể trên màng đặc hiệu với GH (GH receptor (GHR)) ở trên các mô khác nhau, đầu tiên thường là ở mô gan. GH kích thích tế bào sản xuất IGF1, IGF1 được vận chuyển bởi các protein bám với IGF (IGF binding protein (IGFBP)) tới nhiều loại tế bào khác nhau có các thụ thể IGF (IGF receptor (IGFR)). IGFs kích hoạt sự phân bào và các quá trình chuyển hóa khác như chuyển hóa đường, nó được coi là chất trung gian quan trọng giữa GH và sự kích thích tăng trưởng tại cấp độ mô cơ ở động vật có xương sống (hyperplasia và hypertrophy) thông qua sự điều khiển ở gan. Gan là nơi sản xuất IGFBP đóng vai trò tăng cường/ức chế và kéo dài hoạt động của các IGFs (Hình 1.3). Ngoài ra tham gia vào hệ trục này còn có các hormone khác như insulin, leptin, glucocorticoid và

hormone tuyến giáp điều khiển sự tổng hợp và luân chuyển của GH và/hoặc IGF1 [21].

Ngoài hệ trục hormone tăng trưởng, các nhân tố tăng trưởng (Growth Factors (GFs)) (như myostatin – MTTN) và các nhân tố điều hòa cơ (Myogenic Regulatory Factors (MRFs)) (như myogenin (MyoG), MyoD, myf-5 và myf-6) thường được sản xuất ở mô cơ cũng tham gia vào sự tăng trưởng (phát triển cơ) ở cá xương (Hình 1.3). Về tác động qua lại giữa GFs và MRFs, một số nghiên cứu chỉ ra rằng GFs như myostatin có thể là đích gắn trực tiếp của GH ở các sợi cơ và làm giảm biểu hiện của MRFs, vốn chịu trách nhiệm ức chế hoặc tăng cường sự biệt hóa nguyên bào cơ thành sợi cơ, trái ngược với ý kiến cho rằng cấu trúc của MRFs bám vào motif điều hòa ở đầu 5'UTR của GFs làm tăng hoặc giảm sự biểu hiện của GFs. Tuy nhiên, đối với sự tăng trưởng của động vật có xương sống, cả GFs và MRFs đều đóng vai trò tích cực trong quá trình biệt hóa các nguyên bào cơ thành tế bào sợi cơ và được coi là các ứng viên quan trọng trong nghiên cứu tăng trưởng ở cá xương [22].

1.2.3. Một số gen liên quan đến tính trạng tăng trưởng đã được nghiên cứu ở cá xương

Dựa trên vai trò của hệ trục hormone liên quan đến tăng trưởng, các nhân tố tăng trưởng và các yếu tố điều hòa cơ lên sự phát triển cơ - một trong những tính trạng tăng trưởng quan trọng nhất của cá xương, các gen quy định các protein thuộc hệ trục này hoặc các protein điều hòa mức độ phát triển của cơ thường được lựa chọn làm các gen ứng viên để nghiên cứu sự đa hình có liên quan đến các tính trạng tăng trưởng [22].

Nhiều nghiên cứu đã chỉ ra mối liên quan giữa đa hình ở các gen này và tính trạng tăng trưởng của cá xương. Ví dụ như tính đa hình của gen hormone tăng trưởng *GH* (Growth hormone) có liên quan đến tính trạng tăng trưởng được nghiên cứu bằng kỹ thuật giải trình tự trực tiếp trên 282 cá thể cá quế *Siniperca chuatsi* cho thấy: có 4 đa hình đơn nucleotide (SNP) được tìm thấy trên gen *GH* gồm 2 đa hình trên intron 4 (g.4940A>C, g.4948A>T), 1 đa hình trên exon 5 (g.5045T>C) và 1 đa hình trên intron 5 (g.5234T>G). Trong đó, trừ đa hình g.4940A>C, 3 đa hình còn lại đều có tác động đến tính trạng tăng trưởng của cá quế, có thể được sử dụng làm chỉ thị phân tử liên quan đến chọn lọc tính trạng tăng trưởng của loài cá này [23]. Một nghiên cứu mới hơn cũng tìm kiếm mối tương quan giữa sự đa hình chỉ thị phân tử trên gen *GH*

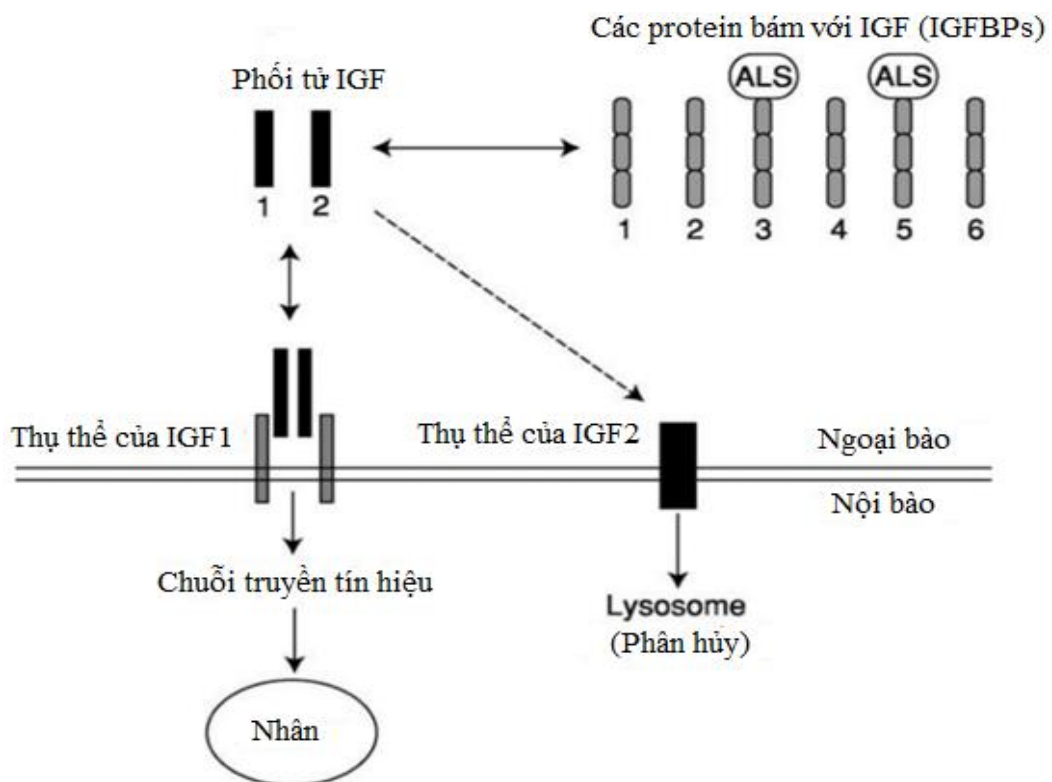
và tính trạng tăng trưởng của *S. chuatsi* bằng cách nhân toàn bộ vùng trình tự của gen *GH*, từ đó xác định được 32 chỉ thị SNP và SSR (Simple sequence repeat- đoạn lặp trình tự đơn giản). Trong số đó, 8 SNP (G1 đến G8) và 1 SSR được lựa chọn để tiến hành genotyping và phân tích mối tương quan với tính trạng tăng trưởng trong một quần thể ngẫu nhiên, từ đó, chỉ ra sự liên quan chặt chẽ giữa các đa hình này với tính trạng tăng trưởng của cá quế. Các phương pháp phân tích tổng hợp nhiều SNP cho thấy các SNP tạo thành 4 kiểu gen đơn bội (haplotype), trong đó, kiểu gen đơn bội thứ nhất thể hiện mối liên quan chặt chẽ với một số tính trạng tăng trưởng hơn cả. Như vậy, các SNP trên gen *GH* của cá quế có thể được coi như những chỉ thị phân tử tiềm năng chọn lọc tính trạng tăng trưởng cho loài cá này [24].

Wenne (2018) đã tổng hợp các kết quả nghiên cứu về ứng dụng dấu chuẩn SNP trong nuôi trồng thủy sản và đánh giá tác động của chúng lên các quần đàn tự nhiên [25]. Theo đó, các nghiên cứu về cá hồi Bắc cực ở Canada *Salvelinus alpinus* đã tìm ra sự tương quan giữa SNP trên gen *GHRH/PACAP2* (gen quy định hormone kích thích tiết hormone tăng trưởng và gen quy định chuỗi polypeptide hoạt hóa enzyme adenylate cyclase của tuyến yên) và sự tăng trưởng ở giai đoạn đầu của loài cá này. Ở cá hồi Đại Tây Dương *Salmo salar*, nghiên cứu trên gen *IGF1* chỉ ra có một SNP trên promoter và một SNP trên intron 3 có mối liên quan chặt chẽ tới khối lượng cơ thể, có tiềm năng trở thành chỉ thị phân tử cho tính trạng tăng trưởng của loài cá này [26]. Đối với cá chép *Cyprinus carpio*, một SNP trên intron 2 của gen *IGF1* cũng có mối liên quan chặt chẽ với tính trạng khối lượng cơ thể và chiều dài thân [27]. Kiểu gen tạo nên từ SNP thuộc vùng 5'-UTR của gen *IGF1* của cá vược châu Âu *Dicentrarchus labrax*, L. được chứng minh là có liên quan đến chiều dài và khối lượng cơ thể của loài cá này [28]. Ở cá chép, 2 SNP ở exon 3 c.42A > G và c.72C > T trên gen *MTTN* có liên quan chặt chẽ tới tính trạng khối lượng cơ thể và các yếu tố điều kiện [29]. Một nghiên cứu khác trên loài cá ngõ *Ancherythroculter nigrocauda* cũng chỉ ra SNP g.1129T>C và g.1289G>A trên gen *MTTN-1* có liên quan đến sự tăng trưởng nhanh của loài này [30]. Ở cá rô phi sông Nin *Oreochromis niloticus*, nghiên cứu SNP trên các gen liên quan đến tăng trưởng (bao gồm: *GH*, *IGF-1* và *MyoG*) đã tìm ra các SNP mới trên gen *GH* và *MyoG* có tiềm năng ứng dụng làm chỉ thị phân tử chọn giống [31].

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

1.3.1. Vai trò, chức năng của một số protein thuộc hệ thống IGF đối với sự tăng trưởng của sinh vật

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 Receptors (IGFRs)) 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 tăng trưởng, biệt hóa và tăng sinh tế bào (Hình 1.4) [32].



Hình 1.4. Hệ thống IGF ở động vật có vú [32]

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. IGFBP-3 và -5 có thể tạo cấu trúc bậc ba với IGFs và các tiểu đơn vị axit không bền (acid-labile subunit (ALS)); phần lớn các IGF trong huyết thanh có thể tạo phức hợp với IGFBP-3 (Hình 1.4).

IGF1 và IGF2 là những chuỗi polypeptide có độ bảo thủ cao, được tổng hợp dưới dạng tiền thân với 5 domain là B, C, A, D, E, sau đó domain E sẽ bị cắt bỏ bởi quá trình phân hủy protein để tạo thành chuỗi peptit IGF trưởng thành. Mức độ tuần hoàn của các IGF trong máu được xác định chủ yếu bởi khả năng tiết IGF của gan, dưới sự điều khiển của hệ trục tuyến yên/hormone tăng trưởng [33]. Các phối tử IGF có nhiều chức năng đa dạng, cần thiết cho sự phát triển bình thường và khả năng sống của động vật thông qua con đường kích thích nội tiết và kích thích tự tiết/ cận tiết [34]. 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 [35].

Sự đáp ứng tế bào đối với IGF1 được điều khiển bởi thụ thể IGF1 (IGF1R) thuộc siêu họ enzyme tyrosine kinase. Cả IGF1 và IGF2 đều có thể bám vào IGF1R, kéo theo sự hoạt hóa của chuỗi dẫn truyền tín hiệu bao gồm con đường PI3/AKT/mTOR (Phosphatidylinositol-3-kinase/ Protein kinase B/ Mammalian target of rapamycin) và MAP (Mitogen-activated protein) kinase, dẫn đến sự phân chia, sống sót và biệt hóa của tế bào [36]. Không giống với thụ thể IGF1R, thụ thể IGF2R có cấu trúc tương tự với thụ thể mannose 6-phosphate và không chứa trung tâm hoạt động của enzyme tyrosine kinase. Mặc dù chuỗi dẫn truyền tín hiệu từ IGF2R được xác định là nguyên nhân phát triển của nguyên bào cơ tim [37], hiện nay chức năng của thụ thể vẫn chưa được làm rõ hoàn toàn. Xét về vai trò đối với hệ thống IGF, thụ thể IGF2R khi liên kết với IGF2 sẽ dẫn đến sự phân hủy các phối tử, làm ổn định nồng độ của phối tử IGF [34].

Hoạt động của các IGF còn được điều khiển bởi các protein bám đặc hiệu IGF (insulin-like growth factor binding proteins (IGFBPs)). Các IGFBPs thuộc họ protein tiết, được chia làm 6 loại từ IGFBP-1 đến -6, có ái lực cao với IGF [38] [39]. Các protein này được đặc trưng bởi cấu trúc rất đặc thù ở đầu amino và đầu carboxyl, do chứa 12 đến 18 phân tử cysteine có độ bảo thủ cao. Các IGFBP khác nhau sẽ được biểu hiện tại các giai đoạn phát triển khác nhau của mô, có tác dụng điều phối hoạt động của các phối tử IGF. Trong đó, IGFBP-1 hạn chế các tín hiệu của IGF khi cơ thể rơi vào các trạng thái dị hóa như đói, stress và giảm oxy huyết, được coi như một nhân tố ức chế tăng trưởng của tế bào sinh dưỡng, dẫn tới quá trình chết của tế bào

[40]. Ở lớp thú, IGFBP-3 tạo phức hợp với ALS bảo vệ các IGF khỏi sự phân hủy của các protease và kéo dài thời gian luân chuyển của IGF trong máu. Ngoài ra, IGFBP-3 còn liên kết với p53 dẫn tới con đường kích thích tự tiết/cận tiết của tế bào, kéo theo các quá trình được điều khiển hoặc phụ thuộc vào các IGF như tăng trưởng, biến nạp và sống sót của tế bào. Tuy nhiên ở lớp cá xương, IGFBP-3 có thể đóng vai trò là chất kích thích hoặc ức chế tăng trưởng, tùy thuộc vào từng loài cá. Mặc dù ít được quan tâm nghiên cứu như IGFBP-1 và IGFBP-3, IGFBP-2 cũng được xác định vừa là tác nhân ức chế cũng là tác nhân kích thích hoạt động của các IGF, đặc biệt là IGF2 và có liên quan đến sự phát triển của hệ thần kinh ở giai đoạn phôi của các loài cá xương [41]. Đối với lớp thú và phần lớn các loài cá xương đã được nghiên cứu, IGFBP-4 đóng vai trò là chất ức chế hoạt động của chuỗi tín hiệu từ IGF, dẫn tới ức chế tăng trưởng, nhưng lại có chức năng kích thích tăng trưởng đối với một số loài cá xương như cá hồi và cá bơn. IGFBP-5 là protein bảo thủ nhất trong họ IGFBP, tương tự với IGFBP-3, protein này cũng liên kết với ALS giúp vận chuyển IGF trong hệ tuần hoàn, nhưng với lượng ít hơn so với IGFBP-3 [42]. Đối với cá xương, IGFBP-5 thúc đẩy quá trình tạo cơ, tăng cơ và hoàn thiện chức năng vây [41]. Riêng với IGFBP-6, mặc dù cơ chế hoạt động chưa được sáng tỏ toàn bộ, nhưng protein này được chứng minh là chất ức chế tăng trưởng trong các điều kiện khẩn cấp. Chính vì các protein trong hệ thống IGF có vai trò quan trọng đối với tăng trưởng, nên việc nghiên cứu đa hình các gen quy định các protein này nhằm tìm ra chỉ thị liên quan đến tính trạng tăng trưởng là rất cần thiết [41].

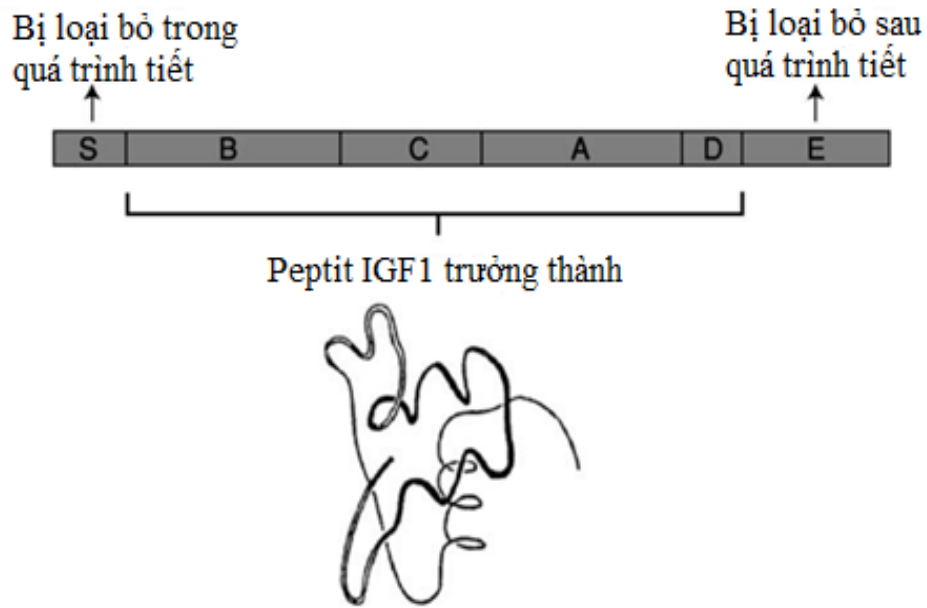
Ngoài các IGFBP-1 đến -6 có ái lực cao với IGF, còn có 9 loại protein liên quan với IGFBP (IGFBP related protein (IGFBP-rPs)) có cấu trúc tương đồng với IGFBP, nhưng có ái lực với IGF thấp hơn các IGFBP [43]. Trong đó, IGFBP-7 hay còn gọi là IGFBP-rP1 bám với IGF1 và IGF2 để điều hòa lượng IGF trong dịch cơ thể và các mô, ngoài ra nó còn có ái lực cao với insulin [44]. Do đó, protein này đóng một vai trò quan trọng trong con đường tín hiệu liên quan đến IGF và insulin (insulin/insulin-like growth factor signaling pathway (IIS)), giúp điều hòa quá trình tăng trưởng, phát triển và sự chuyển hóa năng lượng [45]. Vai trò của IGFBP-7 đến sự tăng trưởng, phát triển của nhiều loài đã được chứng minh. Ở chuột và người, IGFBP-7 được coi như một nhân tố ức chế khối u, sự tăng cường biểu hiện của IGFBP-7 ngăn chặn tín hiệu IGF và sự hình thành mạch, dẫn đến sự ức chế tăng

trưởng tế bào, đồng thời kích hoạt quá trình tự chết tế bào đối với bệnh ung thư đại trực tràng [46], ung thư vú [47], ung thư đường mật [48]. Trong khi đó, sự giảm, mất biểu hiện hoặc sự methyl hóa của gen này có thể dẫn tới bệnh ung thư tuyến giáp [49], ung thư tuyến tiền liệt [50], ung thư tế bào gan [51] và ung thư dạ dày [52]. Trong các quá trình khác liên quan đến tăng trưởng, nguyên bào sợi ở người có thể được biệt hóa thành nguyên bào xương với sự có mặt của IGFBP-7 tái tổ hợp, dẫn tới sự tái tạo xương [53]. Đối với gia súc, gen *IGFBP-7* được biểu hiện mạnh mẽ trong suốt quá trình phát triển nang trứng của trâu [54]. Với sự điều khiển của kích thích tuyến sinh dục, các gen *IGFBP-7* và *IGFBP-6* cùng được tăng cường biểu hiện trong quá trình trưởng thành noãn bào của cá hồi vân *Oncorhynchus mykiss* [55]. Tuy nhiên, hiện có rất ít nghiên cứu về mối liên quan giữa IGFBP-7 và các tính trạng tăng trưởng (chiều dài cơ thể, khối lượng...) của cá xương.

1.3.2. Đặc điểm cấu trúc một số gen và protein tương ứng thuộc hệ thống IGF của cá xương

1.3.2.1. Phối tử IGF (IGF ligands): IGF1 và IGF2

Cũng như các loài động vật có xương sống khác, cDNA của IGF1 của cá xương mã hóa cho dạng tiền thân với chiều dài từ 159 đến 193 axit amin. Tất cả các dạng tiền thân này đều có chuỗi peptit tín hiệu, sau đó sẽ được loại bỏ để tạo thành một tiền phối tử IGF1 với 5 domain theo thứ tự từ đầu N đến đầu C là B-C-A-D-E, tương tự như IGF của lớp thú. Tiếp theo đó, domain E bị loại bỏ để tạo thành IGF1 trưởng thành gồm các domain B-C-A-D với 68 đến 70 axit amin, tùy thuộc từng loài (Hình 1.5). Sự khác nhau về chiều dài này phụ thuộc vào sự có hoặc vắng mặt của 2 axit amin của domain C. Cụ thể, IGF1 trưởng thành của họ cá chép, cá hồi và cá nheo chứa axit amin histidine và asparagine ở vị trí 39 và 40, nếu như tính axit amin đầu tiên của chuỗi peptit trưởng thành là vị trí số 1. Bên cạnh đó, 2 axit amin này lại vắng mặt trong domain C của IGF1 các loài thuộc bộ cá vược, bộ cá mù lùn và bộ cá thân bẹt. Trình tự chuỗi IGF1 trưởng thành của cá xương có 72–81% độ tương đồng với IGF1 trưởng thành ở người và chức năng của protein này cũng được bảo tồn trong suốt quá trình tiến hóa của các loài động vật có xương sống [32].



Hình 1.5. Cấu trúc tiền thân IGF và cấu trúc bậc ba của protein IGF ở cá xương [32].

(Chuỗi peptit tín hiệu (S) bị loại bỏ để tạo thành tiền phối tử bao gồm 5 domain B-C-A-D-E. Domain E tiếp theo đó cũng bị loại bỏ khỏi tiền phối tử để tạo thành IGF1 trưởng thành với 4 domain.)

Có sự khác biệt về cấu trúc gen *IGF1* ở lớp thú và lớp cá. Ở người, IGF1 được mã hóa bởi một gen với 6 exon, trải dài hơn 100 kb trên DNA genome, trong khi đó gen *IGF1* của cá ngựa vằn *Danio rerio*, cá hồi chó *Oncorhynchus keta* và cá bơn vì *Paralichthys olivaceus* chỉ bao gồm 5 exon, trải dài khoảng 15, 22 và 17,5 kb trên genome, tương ứng. IGF1 trưởng thành ở cá hồi được mã hóa bởi exon 2 và 3 trên gen *IGF1*, tương tự như ở lớp thú và gà. Các điểm khởi đầu phiên mã được xác định nằm ở đoạn 385 bp trước codon khởi đầu ATG của gen ở cá hồi chó, cá ngựa vằn và cá rô phi *Oreochromis mossambicus* [32].

Tương tự như IGF1, sau khi loại bỏ chuỗi peptit tín hiệu khỏi dạng tiền thân, tiền phối tử IGF2 cũng chứa 5 domain B-C-A-D-E, tiếp đó, domain E được loại bỏ để tạo IGF2 trưởng thành gồm 67 axit amin ở người, 70 axit amin ở cá xương và 68 axit amin ở cá nhám gai. Sự khác nhau về số lượng axit amin trong chuỗi peptide giữa người và cá xương phụ thuộc vào sự có hoặc vắng mặt các phân tử trong domain C. Ở thú, domain này chứa 8 axit amin, trong khi ở đa phần các loài cá xương có 11 axit amin, ngoại trừ IGF2b ở cá ngựa vằn *D. rerio* và cá nhám gai *Squalus acanthias* chỉ có tương ứng 9 và 8 axit amin trong domain C. Tuy tương đồng về số lượng axit

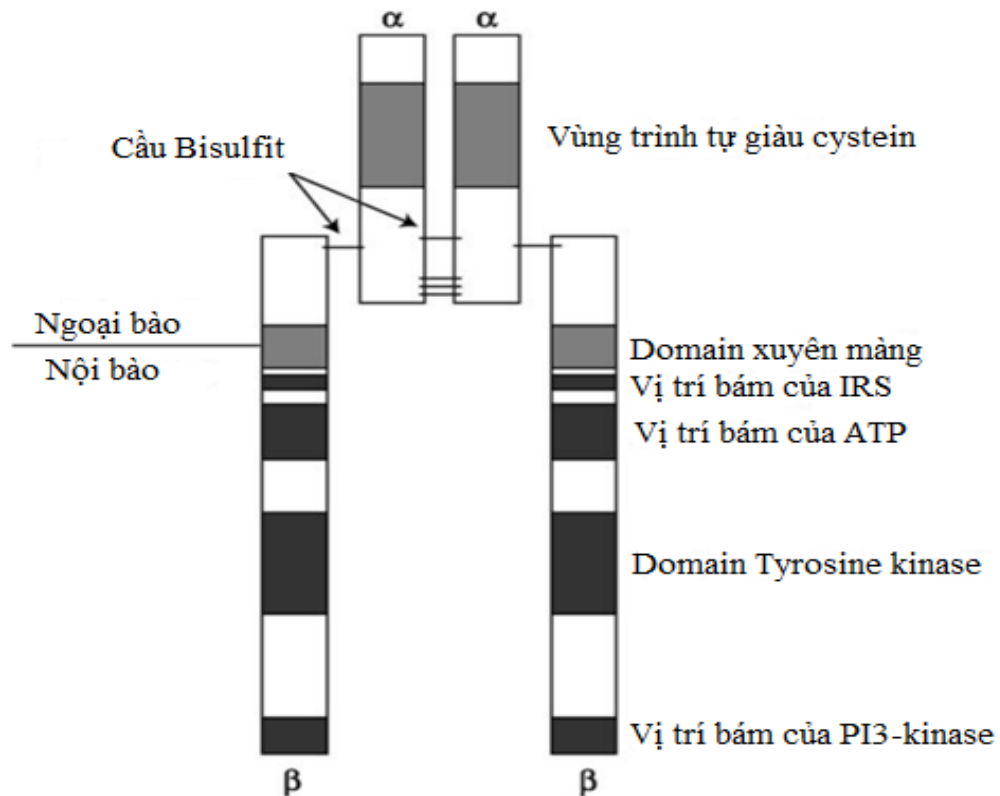
amin trong domain C với lớp thú, trình tự IGF2 của cá nhám gai chỉ tương đồng 66% với trình tự IGF2 của người, trong khi tỉ lệ này giữa IGF2 của người và IGF2 các loài cá xương là 70 đến 75% và giữa các loài cá xương với nhau là 85 đến 100%. Domain A và B của IGF2 có độ bảo thủ cao trong các loài động vật có xương sống với mức độ tương đồng về trình tự đạt từ 70 đến 90%. Năm axit amin đầu tiên của domain B rất khác nhau giữa các loài cá và các loài thuộc lớp thú tuy nhiên chức năng của nó hiện chưa được làm sáng tỏ. Khác với IGF1 có domain D với mức độ đa dạng về trình tự cao, domain D của IGF2 khá bảo thủ giữa các loài cá và các loài động vật có xương sống khác, với 5 trên 6 axit amin tương đồng giữa người và các loài cá xương. Domain E của tiền phối tử IGF2 với 98 axit amin, thường dài hơn so với domain E của IGF1. Một số đặc điểm về cấu trúc IGF2 trưởng thành được bảo tồn trong các loài cá và khác với IGF1, bao gồm: axit amin đầu tiên của domain A luôn là Glu và được bảo tồn ở tất cả các trình tự IGF2 đã được biết, trình tự Glu-Thr-Leu-Cys-Gly (ETLCG) tìm thấy ở domain B được dẫn trước bởi 4 hoặc 5 axit amin trong IGF2, trong khi ở IGF1 là bởi 2 axit amin và domain A của IGF2 chỉ chứa 2 axit amin kỵ nước (RR) trong khi IGF1 lại chứa motif Leu/Phe-Arg-Arg-Leu (L/FRRL) ở domain này [32].

Cấu trúc gen *IGF2* ở 3 loài cá xương gồm cá hồi chó *O. keta*, cá chêm *Lates calcarifer* và họ cá nóc *Tetraodontidae* khi được so sánh với cấu trúc gen *IGF2* của lớp thú cho thấy sự ít phức tạp hơn. Các gen *IGF2* của 3 loài cá này gồm 4 exon và 3 intron, trải dài 7,9 kb, 5,5 kb và 4,2 kb trên genome, tương ứng [32].

1.3.2.2. Thụ thể IGF (IGF Receptor - IGF1R)

Protein kinase (PTK) xúc tác cho quá trình vận chuyển nhóm phosphoryl gamma từ ATP đến tyrosine (tyr) trong cấu trúc protein. IGF1R là một protein thuộc siêu họ protein kinase, đóng vai trò là thụ thể xuyên màng trung gian trong con đường vận chuyển tín hiệu của hệ thống IGF. Protein này được tổng hợp từ một chuỗi peptit đơn, quá trình sau dịch mã sẽ loại bỏ chuỗi peptit tín hiệu dài khoảng 30 axit amin và phân cắt thành hai tiểu phần alpha dài 710 axit amin và hai tiểu phần beta dài 627 axit amin. Tiểu phần alpha và beta được nối với nhau bằng cầu nối bisulfite để tạo nên chuỗi dị dimer anphabeta, hai chuỗi dị dimer anphabeta này tiếp tục được nối với nhau để tạo nên cấu trúc tetramer IGF1R trưởng thành (Hình 1.6). Sự bám của các phối tử như insulin, IGF1, IGF2 vào tiểu phần alpha ngoài màng tế bào của IGF1R sẽ kích hoạt các domain tyrosine kinase ở bên trong tế bào của tiểu phần beta xuyên

màng. Thụ thể được kích hoạt sẽ dẫn tới sự phosphoryl hóa tự động, tăng cường các hoạt động kinase để khởi phát các chuỗi truyền tín hiệu và các chức năng sinh học. IGF1R và Insulin receptor (InsR) có mức độ tương đồng về trình tự ở domain kinase đạt 84% và về chức năng của các domain xúc tác như serine/threonine kinases/ RIO kinases/ aminoglycoside phosphotransferase, choline kinase và phosphoinositide 3-kinase [32].



Hình 1.6. Cấu trúc tetramer IGF1R gồm 2 chuỗi dị dimer alphabeta [32]

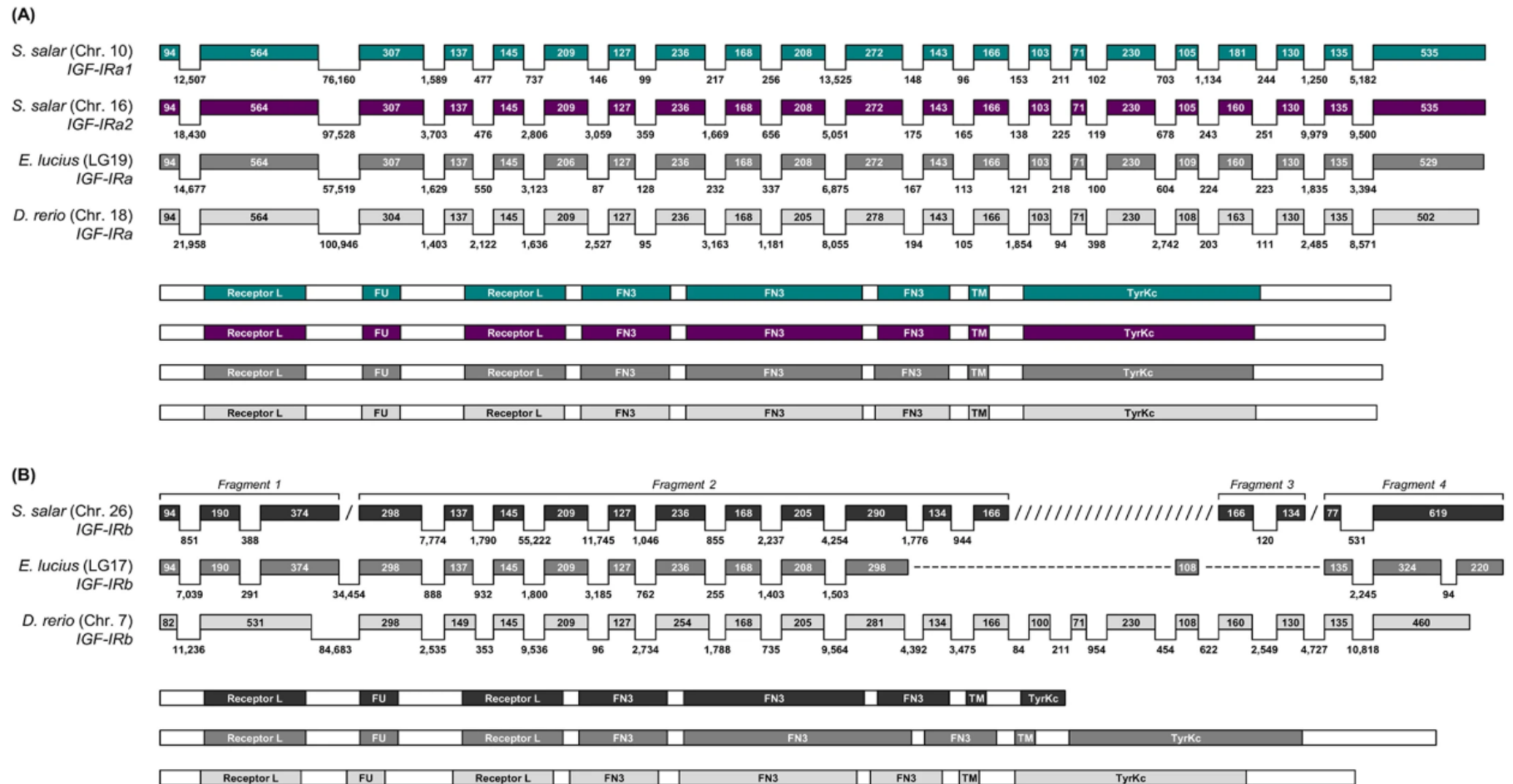
(Hai tiểu phần alpha được liên kết với nhau bằng cầu nối bisulfit, tạo nên vị trí bám với phối tử ở ngoài màng tế bào. Hoạt động tyrosine kinase được xúc tiến bởi tiểu phần beta với cấu trúc xuyên màng. Tiểu phần beta bao gồm các vị trí bám của ATP và PI3-kinase.)

Hai tuýp IGF1R là IGF1Ra và IGF1Rb của cá có mức độ tương đồng cao về cấu trúc, chứa 1405 và 1380 axit amin, tương đồng 63,2 và 59,6% với IGF1R của người, tương ứng. Domain tyrosine kinase ở tiểu phần beta của IGF1R được tìm thấy ở cá ngựa vằn *D. rerio*, cá hồi vân *O. mykiss*, cá thân bẹt *Psetta maxima* và cá bơn vì *P. olivaceus*. Đối với cá ngựa vằn, domain tyrosine kinase ở cả 2 tuýp IGF1R đều chứa một số vùng bảo thủ về chức năng, bao gồm các vị trí bám với ATP, một bộ 3 axit amin tyrosine tương ứng với vị trí phosphoryl hóa tyrosine ở IGF1R giống lớp

thú và motif bám với thụ thể insulin được cho là cần thiết với con đường tín hiệu liên quan đến IGF1R ở lớp thú [32].

Cấu trúc gen *IGF1Ra* và *IGF1Rb* ở các loài cá gồm 21 exon, cách nhau bởi các intron với kích thước đa dạng (Hình 1.7) [56]. Các gen *IGF1Ra* và *IGF1Rb* đều mã hóa cho các domain chức năng từ đầu N đến đầu C lần lượt là domain thụ thể L thứ nhất, domain tương tự Furin giàu cysteine, domain thụ thể L thứ 2, ba domain Fibronectin tuýp 3 (FN3) liên tiếp nhau, domain xuyên màng và domain Tyrosine kinase (Hình 1.7) [56].

Theo Maures và cộng sự (2002), protein IGF1R ở cá ngựa vằn *D. rerio* có chứa vị trí thủy phân gồm 4 axit amin cơ bản là RKRR ở vị trí axit amin thứ 734–737 cho IGF1Ra và RRRR ở vị trí axit amin thứ 727–730 cho IGF1Rb. Domain tyrosine kinase có mức độ bảo thủ cao nhất, bao gồm các đặc điểm cấu trúc liên quan đến chức năng riêng biệt như: một vị trí bám của ATP là GXGXXXG21XK từ axit amin thứ 1004 đến 1033 của IGF1Ra và từ axit amin thứ 994 đến 1021 của IGF1Rb, một cluster chứa 3 tyrosine YXXXYY đóng vai trò chính tạo nên vị trí phosphoryl hóa tự động của IGF1R ở lớp thú, từ axit amin 1159 đến 1164 trên IGF1Ra và từ axit amin 1148 đến 1153 trên IGF1Rb. Motif bảo thủ gắn với cơ chất của thụ thể insulin (insulin receptor substrate-1(IRS-1)-binding motif) NPEY đóng vai trò rất quan trọng trong con đường tín hiệu của IGF1R, lần lượt nằm ở vị trí 975 đến 978 và 965 đến 968 trên IGF1Ra và IGF1Rb. Vùng trình tự giàu cysteine là một nhân tố chính liên quan đến mức độ đặc hiệu của quá trình bám với phối tử của IGF1R, vùng này chứa 26 cystein trải dài trong một khoảng 196 axit amin từ vị trí 156 đến 352 trên IGF1Ra và 151 đến 348 trên IGF1Rb. Ngoại trừ 24 axit amin cuối cùng ở đầu C của IGF1R, các vùng IGF1R nằm trong tế bào chất có mức độ đa dạng cao [57].

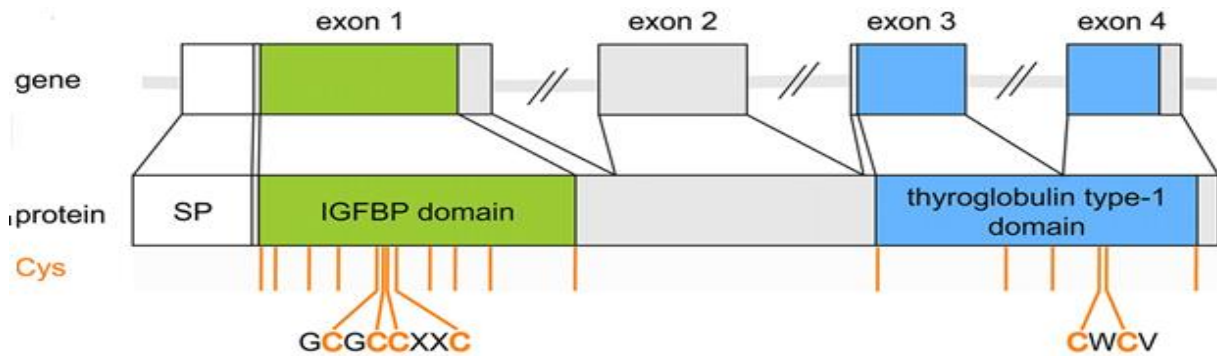


Hình 1.7. Cấu trúc gen, protein IGF1Ra (A) và IGF1Rb (B) của các loài cá [56]

Salmo salar: cá hồi Đại Tây Dương, *Exos lucius*: cá chó phương Bắc, *Danio rerio*: cá ngựa vằn, Receptor L: thụ thể L, FU: Furin-like (tương tự Furin), FN: Fibronectin tuýp III, TM: transmembrane - xuyên màng, TyrKc: Tyrosine kinase.

1.3.2.3. Một số protein bám IGF (IGF binding protein - IGFBP)

Các gen *IGFBP* từ 1 đến 6 đều chứa 4 exon và 3 intron quy định các IGFBP có chiều dài từ 240 đến 328 axit amin và có chung cấu trúc chứa 2 domain bảo thủ về chức năng là domain bám với IGF hay còn gọi là domain IGFBP ở đầu N và domain thyroglobulin tuýp 1 ở đầu C, được nối với nhau bằng domain nối ở giữa. Trong đó, domain IGFBP được mã hóa bởi duy nhất một exon trên gen *IGFBPs*, còn domain thyroglobulin tuýp 1 được mã hóa bởi 2 exon trên gen. Các protein khác liên quan đến họ IGFBP (IGFBP-related protein (IGFBP-rP)) như IGFBP-7 (IGFBP-rP1) cũng chứa domain IGFBP, tuy nhiên domain này chỉ có ái lực yếu với IGF [58].



Hình 1.8. Cấu trúc gen và protein của các IGFBP

Domain IGFBP chứa 60 axit amin với các axit amin cysteine phân bố từ đầu đến cuối domain. Trong tất cả các trình tự IGFBP của động vật có xương sống được khảo sát, domain IGFBP thường chứa 12 phân tử cysteine với độ bảo thủ cao, trừ IGFBP-6 chỉ chứa 10 cystein ở lớp thú và 8 cystein ở lớp cá xương. Trong đó, 4 cystein nằm rải rác trong motif gồm 8 axit amin là Gly-Cys-Gly-Cys-Cys-Xxx-Xxx-Cys (GCGCCXXC), riêng đối với IGFBP-6, chỉ còn 2 cystein và không còn motif này trong domain IGFBP của các loài cá xương [58].

Domain thyroglobulin tuýp 1 ở đầu C của IGFBP thường chứa khoảng 75 axit amin. Mức độ tương đồng về trình tự ở domain này giữa IGFBP-3 và IGFBP-5 đạt cao nhất với 59%, trong khi IGFBP-1 và IGFBP-5 chỉ đạt 31%. Domain này chứa 6 cystein, 2 trong số đó nằm trong motif gồm 4 axit amin là Cys-Trp-Cys-Val (CWCV) được bảo tồn trong tất cả các protein IGFBP của động vật có xương sống cũng như phân ngành Sóng đuôi hay lớp Lưỡng tiêm [58].

Mặc dù có các domain chính tương tự nhau, các IGFBP cũng có một số đặc điểm riêng về cấu trúc. Motif Asn-Xxx-Ser/Thr (NXS/T) trên IGFBP-3 tạo điều kiện để carbohydrate bám vào asparagine. Motif Arg-Gly-Asp (RGD) ở đầu C được bảo tồn trong cấu trúc của IGFBP của các loài động vật có xương sống, bao gồm cả IGFBP-2a và IGFBP-2b của các loài cá xương, IGFBP-3b của cá nóc. Tín hiệu định vị trong nhân bao gồm 10 axit amin Lys và Arg của IGFBP-3 và IGFBP-5 được giữ nguyên ở trình tự của các loài cá xương cũng như các loài động vật có xương sống khác; tuy nhiên số lượng axit amin cơ bản của IGFBP6 của cá xương ít hơn một số với số lượng axit amin cơ bản của IGFBP6 của lớp thú, có thể dẫn đến mức độ tín hiệu định vị nhân thấp hơn [58]. Một motif bám với heparin (PKKXRP) ở IGFBP-2 được coi là bảo tồn ở lớp thú và lớp chim, được giữ lại phần lớn ở cá chêm *L. calcarifer* (PKKTRL), nhưng chỉ còn một phần nhỏ ở cá ngựa vằn *D. rerio* (PK-AP). Chuỗi peptit được mã hóa bởi cDNA IGFBP-1 lần lượt có 40% và 39% độ tương đồng về trình tự với IGFBP-1 và IGFBP-4 của người [59].

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

1.4.1. Vai trò quan trọng của SNP trong việc xác định các chỉ thị phân tử liên quan đến tính trạng quan tâm

Đa hình đơn nucleotide (Single nucleotide polymorphism (SNP)) là một dạng đa hình được tạo ra do sự thay thế, thêm hoặc bớt một nucleotide tại một vị trí đặc biệt trong locus. SNP được chú ý nhất trong các chỉ thị phân tử do nó có thể được xác định một cách tường minh nhất trong genome của bất kỳ loài nào (cả ở vùng mã hóa và không mã hóa), phản ánh sự đa hình tiềm ẩn mà các chỉ thị được phát hiện bằng phương pháp khác không thể hiện được [60], [61], [62]. Trong toàn bộ hệ gen, SNP xuất hiện khá nhiều ở các vùng không mã hóa. Ở vùng mã hóa, SNP có thể dẫn tới sự thay đổi trình tự axit amin (non-synonymous SNPs), hoặc không làm thay đổi trình tự axit amin (silent) tuy nhiên có thể làm thay đổi sự cắt nối mRNA (mRNA splicing).

Cho dù trình tự toàn bộ hệ gen của một số loài thủy sản chưa được biết rõ [63], sự phát triển chỉ thị SNP có thể dựa vào công nghệ giải trình tự gen thế hệ mới (next

generation sequencing (NGS)) thông qua kỹ thuật genotyping toàn bộ các điểm được cắt bởi enzyme cắt giới hạn (restriction site– associated DNA sequencing (RAD-seq)). SNP có thể được sử dụng để lập bản đồ liên kết. Trong thủy sản, SNP thường được sử dụng rộng rãi trong việc xác định các chỉ thị để chẩn đoán bệnh [64], các nghiên cứu liên quan đến các chương trình chọn giống như a) chỉ thị phân tử hỗ trợ chọn lọc dựa trên hệ gen, b) lập bản đồ QTL và nhân dòng vị trí (positional cloning) - xác định các gen liên quan đến tính trạng quan tâm trên NST, c) phân tích kiểu gen đơn bội và phá hệ và d) giám sát hiệu suất của các tổ hợp alen khác nhau trong môi trường đích [65], từ đó tìm kiếm được các SNP tiềm năng có thể được sử dụng làm chỉ thị phân tử.

Đối với những loài đã có genome tham chiếu, trình tự về các gen liên quan đến tính trạng quan tâm có thể được sử dụng để thiết kế các cặp mồi khuếch đại một phần hoặc toàn bộ các gen đó lên. Bằng phương pháp Sanger, các gen ứng viên đó sẽ được giải mã ở nhiều cá thể, từ đó, so sánh trình tự của các cá thể thuộc nhóm đối chứng và các cá thể thuộc nhóm thử nghiệm để xác định được các SNP và tìm ra mối tương quan giữa sự đa hình đó với khác biệt về kiểu hình tính trạng. Quy mô của thí nghiệm kiểm tra sẽ được mở rộng với sự tăng lên về số lượng cá thể trong mỗi nhóm nhằm đảm bảo tính chính xác của xác suất thống kê. Bằng phương pháp này, các SNP trên gen đích được khẳng định tham gia trực tiếp vào điều khiển sự di truyền của tính trạng, hoặc rất gần, có liên kết không cân bằng với chỉ thị quy định tính trạng sẽ được coi là có mối tương quan với tính trạng quan tâm, có thể được sử dụng trong chọn giống [22]. Mặc dù chưa nhiều nhưng một số gen tiềm năng quy định các tính trạng sản lượng, khả năng phòng bệnh hoặc miễn cảm với bệnh đã được phát hiện, tạo tiền đề cho việc xác định sự đa hình chỉ thị phân tử liên quan đến tính trạng mong muốn ở một số loài thủy sản.

1.4.2. Một số phương pháp xác định SNP (SNP genotyping) trên gen đích

1.4.2.1. Phương pháp giải mã lại trên nhiều cá thể

Trong trường hợp chưa biết chính xác số lượng và vị trí các SNP trên gen đích thì việc sử dụng phương pháp giải trình tự Sanger để giải mã lại đoạn gen đích đó trên nhiều cá thể và so sánh để tìm kiếm SNP sẽ mang lại hiệu quả tốt. Để tiến hành phương pháp này, các đoạn gen đích cần được nhân lên một cách đặc hiệu bằng phản ứng PCR, sau đó được sử dụng làm khuôn cho phản ứng giải trình tự toàn bộ đoạn gen đích đó. Ưu

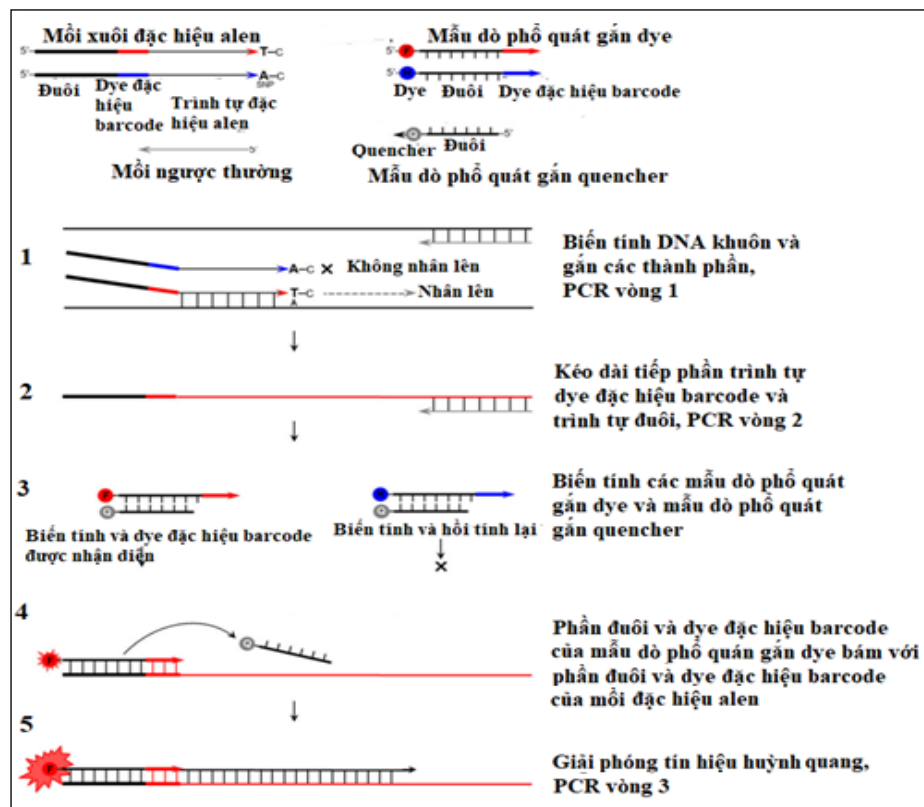
điểm của phương pháp này là có độ chính xác, đặc hiệu cao, xác định được đồng thời nhiều SNP. Nhược điểm của phương pháp này là đòi hỏi kỹ thuật cao, cần nhiều thời gian, chi phí và trang thiết bị để tiến hành các thí nghiệm PCR đặc hiệu và giải trình tự. Do đó, SNP genotyping bằng phương pháp giải mã lại trên nhiều cá thể sẽ phù hợp với yêu cầu xác định SNP trên gen đích với độ chính xác cao nhưng ở số lượng cá thể nhỏ.

1.4.2.2. Phương pháp lai phân tử

Phương pháp này sẽ được áp dụng khi SNP cần được xác định đã được biết rõ thông tin về vị trí trên gen đích và các alen có thể xuất hiện tại vị trí SNP đó (allele discrimination). Với mỗi vị trí SNP thuộc DNA đích cần được xác định, một cặp gồm hai trình tự đầu dò đặc hiệu alen sẽ được thiết kế để lai với DNA đích chỉ khi sự bắt cặp bổ sung diễn ra hoàn toàn. Trong điều kiện thí nghiệm được tối ưu hóa, chỉ cần tại vị trí SNP không có sự bắt cặp giữa đầu dò và DNA đích cũng khiến cho phản ứng lai bị ngăn chặn. Cụ thể, các đầu dò đặc hiệu alen được cố định trên bề mặt rắn, trong khi các đoạn DNA cần giải trình tự SNP đã được dán nhãn tín hiệu không được cố định. Nếu sự bắt cặp hoàn toàn giữa đầu dò và đoạn DNA diễn ra, đoạn DNA sẽ được bắt giữ lại trên bề mặt, khi đó nhãn tín hiệu của đoạn DNA được nhận diện cho thấy phản ứng lai đã diễn ra. Ngược lại, nếu không có sự bắt cặp hoàn toàn, đoạn DNA sẽ không lai được với đầu dò và bị rửa trôi, dẫn tới không có tín hiệu được nhận diện. Do đó, khi biết chính xác vị trí của các đầu dò được cố định trên bề mặt rắn, qua phản ứng lai phát/không phát tín hiệu, kiểu gen của đoạn DNA liên quan đến SNP sẽ được làm rõ [66]. Một bề mặt được cố định nhiều cặp đầu dò đã biết rõ vị trí sẽ giúp xác định nhiều vị trí SNP khác nhau trên DNA đích cùng một lúc, đây chính là cơ sở của phương pháp SNP array. Phương pháp lai có nguyên lý đơn giản nhất trong các phương pháp xác định SNP do không sử dụng enzyme trong quá trình thực hiện, nên dễ tiến hành trong thời gian ngắn, không đòi hỏi thao tác thí nghiệm phức tạp. Thách thức lớn nhất của phương pháp này nằm ở việc thiết kế đầu dò thật sự đặc hiệu, tuy nhiên với sự trợ giúp của các phần mềm thiết kế đầu dò ngày càng được tối ưu hóa thì tỉ lệ thiết kế đầu dò thành công cho phản ứng lai ngày càng cao [66].

1.4.2.3. Phương pháp PCR đặc hiệu alen (Allele-specific q-PCR) dựa trên FRET (ASQ)

Trong phương pháp PCR đặc hiệu alen dựa trên hệ chuyển đổi năng lượng cộng hưởng huỳnh quang (Fluorescence resonance energy transfer (FRET)) (ASQ), huỳnh quang và quencher bắt giữ được gắn vào hai oligonucleotide riêng biệt có trình tự bổ sung với nhau. Hai thành phần được đưa vào đồng thời trong cùng một phản ứng PCR gồm hỗn hợp các môi đặc hiệu alen và môi thường; và hỗn hợp phát hiện không phụ thuộc khuôn DNA bao gồm 2 đến 4 mẫu dò phổ quát (UP-1 đến -4) và một mẫu dò oligonucleotide phổ quát cho quencher (Uni-Q). Vị trí SNP được định vị tốt hơn ở vị trí áp chót của mỗi môi đặc hiệu alen, giúp tăng tính đặc hiệu của phản ứng và khả năng phân biệt alen (Hình 1.9).

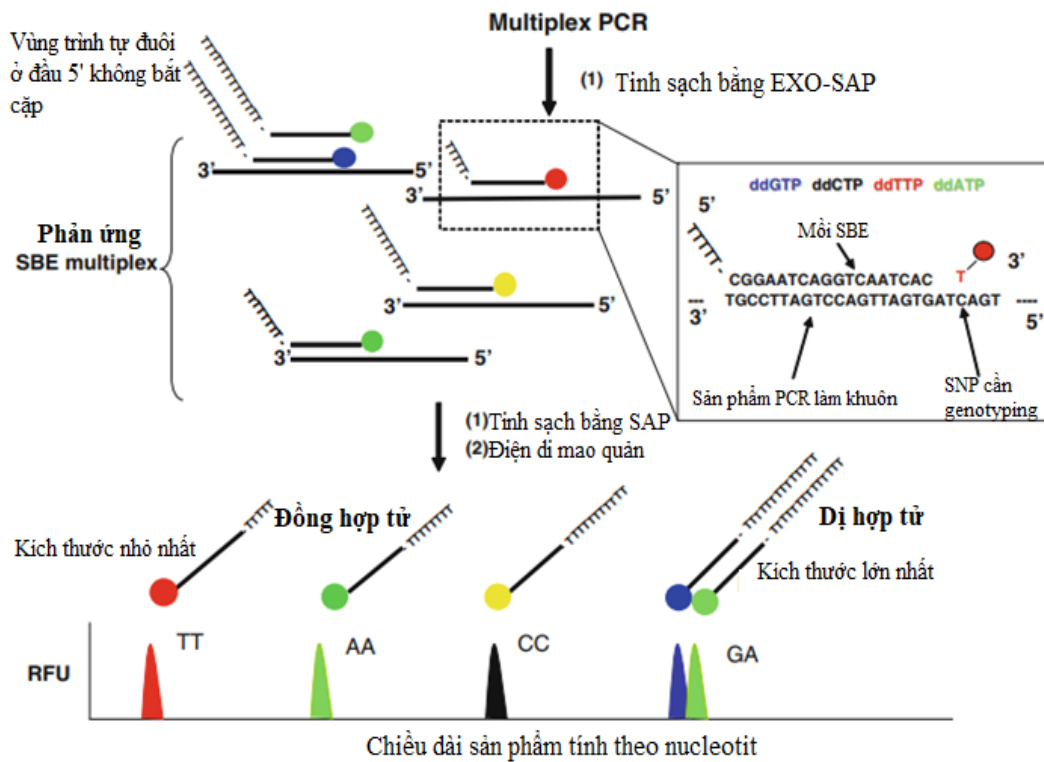


Hình 1.9. Nguyên lý kỹ thuật PCR đặc hiệu alen dựa trên FRET

Ưu điểm của phương pháp này là có thể xác định rõ ràng được lượng huỳnh quang phát ra, ít bị nhiễu tín hiệu, quy trình đơn giản, dễ thực hiện, giá thành rẻ hơn nhờ các thành phần phổ quát [67].

1.4.2.4. Phương pháp kéo dài một nucleotide (Single base extention)

Phương pháp kéo dài một nucleotide (Single base extention (SBE)) được tiến hành trong điều kiện các SNP cần xác định đã được biết rõ vị trí trên gen đích, tuy nhiên không cần phải biết rõ các alen có thể có ở vị trí SNP đó. Nguyên lý của phương pháp này là mỗi SBE sẽ được thiết kế đặc hiệu ngay trước vị trí đầu 3' của SNP cần xác định trên gen đích; thông qua phản ứng PCR, sau khi mỗi SBE được gắn vào khuôn, 1 ddNTP có gắn huỳnh quang phân biệt cho các nucleotide A, T, G, C sẽ được bắt cặp bổ sung với khuôn tại chính vị trí SNP cần xác định và phản ứng dừng lại, do đó phương pháp này có tên là kéo dài một nucleotide. Thông qua điện di mao quản sản phẩm và đọc tín hiệu huỳnh quang của một ddTNP được gắn thêm trên máy giải trình tự (mini-sequencing), trình tự SNP sẽ được xác định [68].



Hình 1.10. Hình ảnh minh họa quá trình thực hiện thí nghiệm SBE sử dụng SNaPshot

Phương pháp này cho phép phát hiện đồng thời đến hơn 30 điểm SNPs nằm rải rác trên gen đích [69]. Ưu điểm của phương pháp này là khả năng phát hiện tetra – allelic SNPs, nhanh nhạy, có tính đặc hiệu và khả năng giải trình tự tự động. SBE đã được áp dụng trong một vài ứng dụng khác như phân tích tế bào đơn dòng cho chẩn đoán di

truyền trước khi cấy ghép, chẩn đoán phân tử trước và sau sinh [70], phân tích và giám định DNA ti thể trên mẫu đã bị phân hủy [71] và sàng lọc SNP hiệu năng cao trong nghiên cứu quần thể [72]. Nhược điểm của phương pháp này là đòi hỏi kỹ thuật, chi phí cao và trang thiết bị phức tạp để thực hiện.

Một trong những công nghệ đã được thương mại hóa trên cơ sở kỹ thuật của phương pháp SBE là bộ kit thương mại SNaPshot Multiplex Kit của hãng Applied Biosystems, Mỹ. Phản ứng kéo dài một nucleotide trong bộ kit SNaPshot sử dụng các ddNTP được đánh dấu huỳnh quang khác nhau, gồm dATP màu xanh lá (dR6G), dTTP màu đỏ (dROX), dCTP màu đen (dTAMRA) và dGTP màu xanh dương (dR110). Để thực hiện phản ứng multiplex xác định nhiều SNP trên các đoạn gen đích, các mồi SBE được thiết kế gồm phần trình tự lõi bắt cặp bổ sung với trình tự đích tại vị trí 3' nằm ngay sát vị trí SNP và phần trình tự đuôi là các đoạn lặp ở phía 5' của mồi SBE, có chiều dài khác nhau từ 5 đến 8 nucleotide giữa các mồi, nhằm phân biệt về chiều dài các sản phẩm thu được trong quá trình điện di mao quản tự động (Hình 1.10). Theo một số nghiên cứu trước đó, phương pháp kiểm nghiệm chỉ thị SNP bằng bộ kit SNaPshot thường được phân tích bằng phần mềm GeneMapper (ABI) với các phiên bản khác nhau [73]. Phần mềm này khá đa dạng về chức năng như phân tích tính đa hình chiều dài các đoạn khuếch đại (AFLP), xác định trạng thái dị hợp tử của gene đột biến (LOH), microsatellite và phân tích kết quả SNP genotyping.

1.4.2.5. Phương pháp cắt enzyme giới hạn (Enzymatic cleavage)

Enzym đặc hiệu cấu trúc sẽ cắt phức hợp được tạo ra bởi các đầu dò oligonucleotide chồng lấp lên nhau (overlap) hoàn toàn tại vị trí SNP cần xác định. Các đầu dò được thiết kế bao gồm: đầu dò xâm lấn (Invader oligo) bắt cặp với trình tự ở đầu 3' của DNA đích cho đến vị trí liền trước đầu 3' của SNP cần xác định và đầu dò đặc hiệu alen bắt cặp với trình tự từ vị trí SNP cần xác định cho đến đầu 5' của DNA đích. Đầu dò đặc hiệu alen ngoài phần bắt cặp với trình tự đích còn có trình tự đuôi ở đầu 5' được treo lỏng lẻo (Flap) [66]. Ở bước đầu tiên khi đầu dò xâm lấn bắt cặp bổ sung với trình tự đích, nếu đầu dò đặc hiệu alen cũng bắt cặp được hoàn toàn với trình tự đích, phức hợp với các oligonucleotide chồng lấp hoàn toàn sẽ được tạo ra, kích hoạt enzym đặc hiệu cấu trúc cắt rời trình tự đuôi (Flap) và nucleotide ở vị trí chồng lấp ra khỏi đầu

dò đặc hiệu alen. Ở bước thứ hai, trình tự được giải phóng sẽ tiếp tục bắt cặp hoàn toàn với trình tự loop đã chứa sẵn tín hiệu huỳnh quang và quencher trong hệ chuyển đổi năng lượng cộng hưởng huỳnh quang thứ nhất, tạo nên một phức hợp chồng lấp mới, từ đó kích hoạt enzyme cắt tại vị trí chồng lấp, tách tín hiệu huỳnh quang xa khỏi quencher bắt giữ, giải phóng tín hiệu huỳnh quang. Trong trường hợp đầu dò đặc hiệu alen không bắt cặp được hoàn toàn với trình tự đích, phức hợp với các oligonucleotide chồng lấp hoàn toàn sẽ không được tạo ra, không kích hoạt được sự cắt rời trình tự đuôi (Flap), dẫn đến không kích hoạt được hệ chuyển đổi năng lượng cộng hưởng huỳnh quang [66].

Có thể áp dụng phương pháp này để xác định nhiều vị trí SNP trên cùng DNA đích trong cùng một lúc, với tín hiệu huỳnh quang phát ra mạnh mẽ mà không cần phải thông qua phản ứng PCR khuếch đại DNA đích. Tuy nhiên phương pháp này đòi hỏi lượng DNA đích đầu vào phải đủ lớn, các đầu dò phải đặc hiệu và có độ tinh sạch cao, hơn nữa việc thiết kế đầu dò là khá phức tạp vì phải đảm bảo tiêu chí để hai bước phản ứng diễn ra tuần tự trong khi tất cả các thành phần phản ứng của hai bước này được trộn cùng một lúc [66].

1.4.3. Các nghiên cứu về SNP trên một số gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng ở cá xương

Ở cá hồi Đại Tây dương *S. salar* L., từ 16 cá thể, 3 SNP trên gen IGF1 đã được phát hiện gồm 1 SNP trên promoter *g.5763G>T*, 1 SNP trên intron 1 *g.7292C>T* và 1 SNP trên intron 3 *g.4671A>C*. Kết quả genotyping 3 SNP này trên 4800 cá thể ngẫu nhiên đã chỉ 2 SNP *g.5763G>T* và *g.4671A>C* có liên quan chặt chẽ với khối lượng cơ thể và đây cũng là những chỉ thị tiềm năng có thể ứng dụng cho chọn giống [26]. Trình tự gen *IGF1* của 17 cá thể cá chép *Cyprinus carpio* được so sánh với nhau, tìm ra 3 SNP trên intron và 1 SNP trên exon của gen, 4 SNP này sau đó được genotyping trên 289 cá thể được chọn lọc ngẫu nhiên và cho kết quả SNP nằm trên intron 2 của gen *IGF1* (*g.7627T>A*) có liên quan chặt chẽ đến khối lượng và chiều dài cơ thể. Cụ thể, cá chép có kiểu gen AA có khối lượng trung bình cao hơn cá có kiểu gen TT 5,9%, cá có kiểu gen TT tương ứng có khối lượng trung bình thấp nhất [27]. Kết quả phát hiện và kiểm nghiệm SNP liên quan đến tính trạng tăng trưởng trên 80 cá thể cá vược châu Âu *D. labrax*, L. cho thấy kiểu gen tạo nên từ SNP *g.46672A > G* và *g.46749C > T* thuộc vùng

5'-UTR của gen *IGF1* được chứng minh có liên quan đến chiều dài cơ thể của loài cá này, ngoài ra, kiểu gen được tạo nên từ SNP g.46749C > T còn có liên quan đến tính trạng khối lượng cơ thể [28].

Ngoài mức độ biểu hiện của gen *IGF2* được chứng minh có liên quan đến tính trạng tăng trưởng của nhiều loài cá [74], [75], [76], [77], cho đến nay, một số nghiên cứu cũng đã chỉ ra được mối liên quan giữa các SNP trên gen này với tính trạng tăng trưởng của các loài cá xương. Ví dụ ở cá rô phi sông Nin *Oreochromis niloticus* L, thông qua các bước giải trình tự toàn bộ gen *IGF2* trên 10 cá thể, Juhua và cộng sự (2010) đã phát hiện được 13 SNP gồm 9 SNP trên intron và 4 SNP thuộc vùng mã hóa nhưng không làm thay đổi trình tự axit amin, sau đó 4 SNP này được genotyping trên 192 cá thể và kết quả đã chỉ ra SNP G161A trên exon 3 liên quan đến tính trạng tăng trưởng của loài cá này [78]. SNP trên intron 3 của gen *IGF2* được phát hiện sau khi so sánh trình tự của 16 cá thể cá vược châu Âu *D. labrax*, L. và được chứng minh có liên quan đến tính trạng khối lượng cơ thể của loài cá này sau khi SNP được genotyping trên 298 cá thể được chọn lọc ngẫu nhiên [79].

Nghiên cứu phát hiện và kiểm nghiệm SNP trên 195 cá thể cá bản địa Trung Quốc *Odontobutis potamophila* đã phát hiện được 3 SNP trên intron của gen *IGF1R* là 1777T>C, 1704G>A, and 1208G>A, tuy nhiên chỉ có SNP 1208G>A liên quan chặt chẽ đến các tính trạng tăng trưởng quan trọng. Các cá thể có kiểu gen AG có sự phát triển tốt hơn về chiều dài thân, chiều rộng thân, khối lượng cơ thể so với các cá thể có kiểu gen GG [80]. Tuy vậy, các nghiên cứu về mức độ biểu hiện gen và đa hình phân tử, đặc biệt là SNP trên gen *IGF1R* liên quan đến tính trạng tăng trưởng ở các loài cá còn rất hiếm.

Haplotype, tổ hợp kiểu gen phát sinh từ SNP trên các gen thuộc hệ thống IGF cũng được chứng minh có liên quan đến tăng trưởng của cá xương. Haplotype tạo nên từ 3 SNP trên gen *IGF1* liên quan chặt chẽ với khối lượng cơ thể lúc thu hoạch và tỉ lệ phi lê của cá hồi Đại Tây Dương *S. salar* [26]. Tổ hợp kiểu gen tạo nên từ 2 SNP trên gen *IGF1* cũng được chứng minh có liên quan đến tính trạng chiều dài cơ thể và khối lượng cơ thể của cá chép *C. carpio* [27].

1.4.4. Các gen thuộc hệ thống IGF đã được xác định dựa trên thông tin hệ gen cá tra nuôi

Sử dụng kỹ thuật giải trình tự thế hệ mới (hệ thống Illumina), toàn bộ hệ gen (genome) của một cá thể cá tra đực đã được giải mã bao gồm dữ liệu thư viện pair-end và dữ liệu từ thư viện mate pair. Kết hợp với dữ liệu transcriptome từ các mô cơ quan và các giai đoạn phát triển khác nhau, genome cá tra đã được dự đoán mô hình gen và chú giải, dự đoán có 28600 gen mã hóa protein. Tất cả các thông tin về trình tự genome của cá tra nuôi được công bố trên ngân hàng dữ liệu DDBJ/EMBL/NCBI với mã số BioProject là PRJNA448819, hệ gen lắp ráp được công bố lần đầu tiên với mã số QUXB00000000. Bộ dữ liệu này cũng được công bố trên genome browser tại địa chỉ <http://marinegenomics.oist.jp/gallery/> hoặc <http://catfish.genome.ac.vn>. Đây là dữ liệu genome đầu tiên của cá tra, có thể dùng làm trình tự tham chiếu cho các nghiên cứu sâu sau này [16].

Dữ liệu hệ gen lắp ráp của cá tra nuôi đã được công bố trên ngân hàng dữ liệu, từ đó cung cấp các thông tin về mã số định danh của các scaffold, gen, mRNA và protein cho người sử dụng. Theo đó, các gen thuộc hệ thống IGF của cá tra nuôi đã được xác định (Bảng 1.1) [16]. Cụ thể, trong genome của cá tra nuôi đã xác định được 2 gen quy định cho IGF1 (IGF1 và IGF isoform X1) và 2 gen quy định cho IGF2, tương tự với 4 gen quy định IGF1a, 1b, 2a, 2b ở cá ngựa vằn *D. rerio* và nằm trên các scaffold khác nhau. Trong khi ở cá ngựa vằn có hai gen quy định thụ thể IGF – IGFR (IGF1Ra và IGF1Rb) nằm trên nhiễm sắc thể số 2 và 22, thì ở cá tra nuôi, có 3 gen mã hóa cho thụ thể IGF (IGFR, IGF1R, IGF1R isoform X1). Thêm vào đó, 11 gen quy định IGFBP gồm IGFBP-1, -2A, -2B, -3, -5, -6, -7 được tìm thấy trong genome của cá tra nuôi, tuy nhiên không có gen nào quy định IGFBP-4, khá tương đồng với cá ngựa vằn (Bảng 1.1). Căn cứ vào vai trò của một số gen thuộc hệ thống IGF liên quan đến quá trình tăng trưởng của các loài động vật có xương sống nói chung và các loài cá xương nói riêng, vào các cơ sở dữ liệu về trình tự genome tham chiếu của cá tra nuôi *P. hypophthalmus* với độ tin cậy cao, cũng như vào tính khả thi của phương pháp nghiên cứu đa dạng di truyền (SNP) trên các gen đích liên quan đến tính trạng cần quan tâm, trong đề tài này, một số gen thuộc hệ thống IGF của cá tra nuôi (Bảng 1.2) sẽ được phân tích, phát hiện các đa hình

(SNP) của các gen đó và kiểm nghiệm sự liên quan của chúng với tính trạng tăng trưởng trong quần thể cá tra tăng trưởng nhanh và tăng trưởng chậm, hướng tới tìm kiếm chỉ thị phân tử liên quan đến tính trạng tăng trưởng

Bảng 1.1. Các gen thuộc hệ thống IGF của cá tra nuôi

Họ gen	Số scaffold	Gene IDs	Mô tả	Chiều dài nucleotide	Chiều dài amino axit
<i>IGF</i>	11	phy_g435.t1	insulin-like growth factor I	736	245
<i>IGF</i>	16	phy_g25602.t1	insulin-like growth factor I isoform X1	549	183
<i>IGF</i>	28	phy_g10034.t1	insulin-like growth factor II	639	213
<i>IGF</i>	8	phy_g13975.t1	insulin-like growth factor II	711	237
<i>IGFBP</i>	58	phy_g11974.t1	insulin-like growth factor-binding 1	735	245
<i>IGFBP</i>	42	phy_g24144.t1	insulin-like growth factor-binding 1	796	265
<i>IGFBP</i>	46	phy_g8897.t1	insulin-like growth factor-binding 2A	786	262
<i>IGFBP</i>	3*	phy_g8121.t1	insulin-like growth factor-binding 2B	675	225
<i>IGFBP</i>	58	phy_g11973.t1	insulin-like growth factor-binding 3	856	285
<i>IGFBP</i>	42	phy_g24145.t1	insulin-like growth factor-binding 3	906	302
<i>IGFBP</i>	3*	phy_g8120.t1	insulin-like growth factor-binding 5	919	306
<i>IGFBP</i>	46	phy_g8896.t1	insulin-like growth factor-binding 5	760	253
<i>IGFBP</i>	9*	phy_g1604.t1	insulin-like growth factor-binding 6	612	204
<i>IGFBP</i>	22*	phy_g27470.t1	insulin-like growth factor-binding 6	579	193
<i>IGFBP</i>	54	phy_g24954.t1	insulin-like growth factor-binding 7	792	264
<i>IGFR</i>	19	phy_g25100.t1	insulin receptor-like	4026	1342
<i>IGFR</i>	19	phy_g25233.t1	insulin-like growth factor 1 receptor	4057	1352
<i>IGFR</i>	15	phy_g24431.t1	insulin-like growth factor 1 receptor isoform X1	4237	1412
* Hox cluster containing scaffolds					

Bảng 1.2. Số định danh trình tự gen, mRNA, protein trên NCBI của một số gen trong hệ thống IGF của cá tra nuôi được lựa chọn để phân tích cấu trúc

STT	Tên gen (Gene ID ban đầu)	Gene ID	Vị trí trên genome	mRNA ID	Protein ID
1	<i>IGF1</i> (<i>phy_g435.t1</i>)	113532358	sc0000011 NW_020824206.1 (8211039..8230392)	XM_026923725.1	XP_026779526.1
2	<i>IGF2</i> (<i>phy_g10034.t1</i>)	113540971	sc0000028 NW_020824223 (1585727..1590597)	XM_026937691.1	XP_026793492.1
3	<i>IGF1R isoform X1</i> (<i>phy_g24431.t1</i>)	113534424	sc0000015 NW_020824210.1 (3316677..3410561)	XM_026927204.1	XP_026783005.1
4	<i>IGFBP-1</i> (<i>phy_g24144.t1</i>)	113545355	sc0000042 NW_020824237.1 (1915945..1918707)	XM_026944681.1	XP_026800482.1
5	<i>IGFBP-2A</i> (<i>phy_g8897.t1</i>)	113545824	sc0000046 NW_020824241.1 (4399897..4427991)	XM_026945448.1	XP_026801249
6	<i>IGFBP-3</i> (<i>phy_g24145.t1</i>)	113545296	sc0000042 NW_020824237.1 (1921250..1939687)	XM_026944580.1	XP_026800381.1
7	<i>IGFBP-5</i> (<i>phy_g8120.t1</i>)	113525861	sc0000003 NW_020824198.1 (9142631..9154838)	XM_026912477.1	XP_026768278.1
8	<i>IGFBP-6</i> (<i>phy_g1604.t1</i>)	113530376	sc0000009 NW_020824204.1 (3855817..3860008)	XM_026920287.1	XP_026776088.1
9	<i>IGFBP-7</i> (<i>IGFBP-rP1</i>) (<i>phy_g24954.t1</i>)	113546920	sc0000054 NW_020824249.1 (1267971..1277002)	XM_026947045.1	XP_026802846.1

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 [15]. Cụ thể, các cá thể cá tra nuôi được thu thập từ quần đàn chọn lọc theo hướng tăng trưởng dựa trên lý thuyết di truyền số lượng, thuộc chương trình chọn giống tiến hành tại Viện Nuôi trồng Thủy sản 2 (Việt Nam). Chương trình chọn giống này bắt đầu từ năm 2001, tính đến thời điểm thu thập mẫu, chương trình đã khởi tạo 3 thể hệ quần đàn được chọn lọc theo hướng sinh trưởng nhanh là G1, G2 và G3. Các cá thể cá tra phục vụ nghiên cứu này thuộc quần đàn G3-cộng gộp (2014) được phối từ cá bố, mẹ thuộc các quần đàn G3-2001, G2-2002 và G2-2003 theo phương pháp lai tổ hợp thứ bậc. Tổng cộng 20.027 cá thể thuộc 226 gia đình G3-cộng gộp sau khi được ương nuôi đến mức khối lượng nhất định đã được chọn lọc ngẫu nhiên để gắn dấu từ nhận diện PIT (Passive Integrated Transponder), và tiếp tục được nuôi riêng rẽ trong 192 ngày để cân đo khối lượng cơ thể và tính toán giá trị chọn giống ước đoán (Estimated breeding value – EBV). Từ đó, 160 cá thể thuộc 226 gia đình của quần đàn G3-cộng gộp đã được chọn lọc để thu mẫu phục vụ nghiên cứu của luận án vào năm 2015. Mẫu thu được để phục vụ 2 giai đoạn chính của thí nghiệm là phát hiện, sàng lọc SNP trên các gen đích và kiểm nghiệm sự liên quan của SNP lên tính trạng tăng trưởng. tiến hành từ năm 2018 đến năm 2020.

Đầu tiên, SNP trên các gen thuộc hệ thống IGF sẽ được tìm kiếm và sàng lọc trên bộ mẫu khởi tạo. Dựa vào giá trị giá trị chọn giống ước đoán EBV trung bình của cá thể và EBV trung bình của gia đình trên toàn đàn, 10 cá thể có EBV cao nhất thuộc 9 gia đình có EBV trung bình cao nhất được lựa chọn làm 10 mẫu tăng trưởng nhanh (ký hiệu từ N1 đến N10) và 10 cá thể có EBV thấp nhất thuộc 9 gia đình có EBV trung bình thấp nhất được lựa chọn làm 10 mẫu tăng trưởng chậm (ký hiệu từ C1 đến C10). Luận án lựa chọn số lượng 20 cá thể cho bộ mẫu này do có tham khảo các công trình khoa học trên các loài cá hồi Đại Tây dương *S. salar* L. [26], cá chép *C. carpio* [27], cá rô phi sông Nin *O. niloticus* L [78], cá vược châu Âu *D. labrax*, L. [79] đã sử dụng từ 10 đến gần 20 cá thể để phát hiện SNP trên các gen thuộc hệ thống IGF. Đây cũng là số lượng cá thể phù

hợp với quy mô thời gian và chi phí tiến hành thí nghiệm của Luận án. Thông tin về khối lượng cơ thể và EBV của các cá thể được trình bày trong Bảng 2.1.

Bảng 2.1. Thông tin 20 cá thể cá tra nuôi trong bộ mẫu khởi tạo

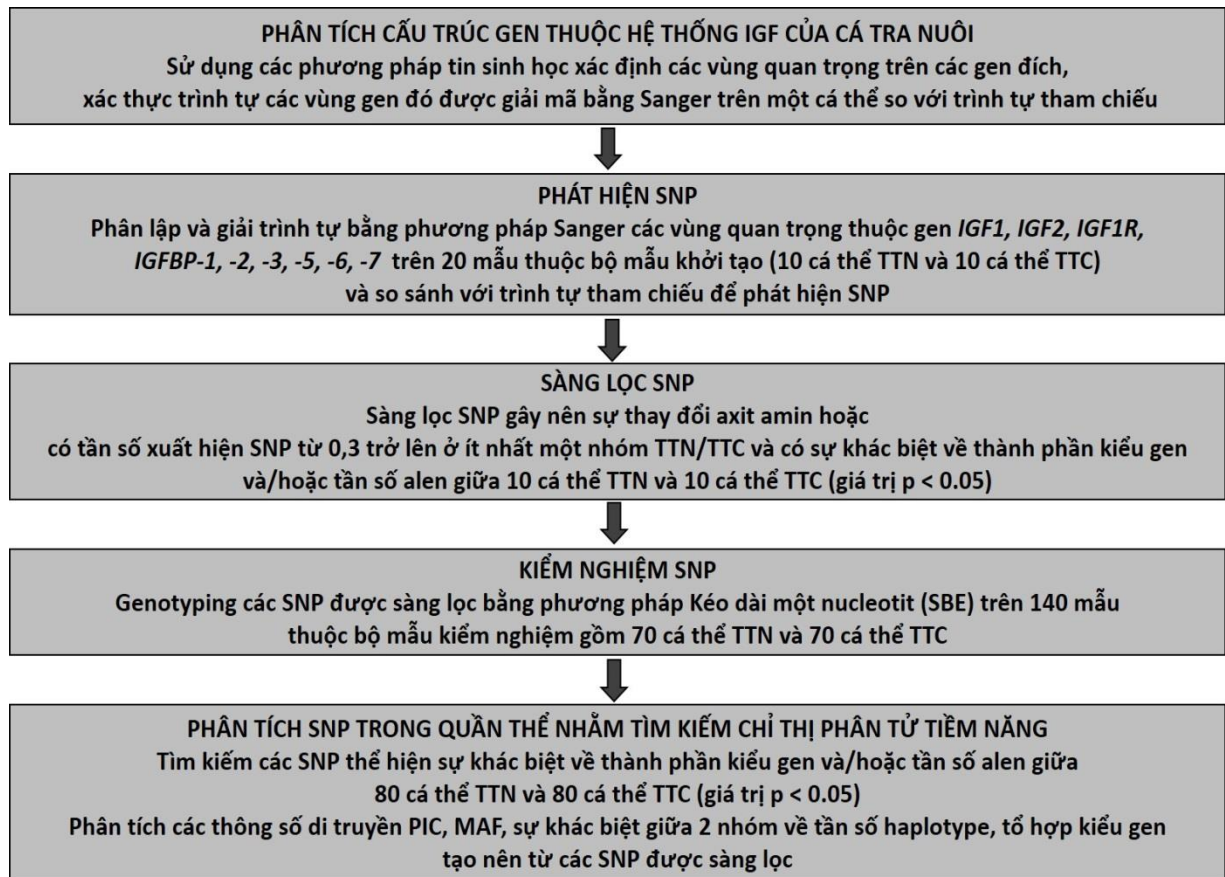
Thông tin 10 mẫu cá thể tăng trưởng nhanh						
STT	ID	Ngày tuổi	Khối lượng (g)	EBV	Gia đình	EBV gia đình
N1	0418CBE232	385	1969,8	400,9	131	173,1
N2	0418CBCBD7	394	2160	439,4	113	251,4
N3	0418CC3BB3	383	1702,6	420,9	38	170,7
N4	0418CBE610	361	1750,4	313,6	161	179,4
N5	0418CBCB67	371	1691,4	314,1	161	179,4
N6	0418CBF2A2	380	2017,8	366,5	175	232,1
N7	0418CBFB40	362	1439,2	310,7	180	213,1
N8	0418CBBF81	406	1585	405,4	30	263,6
N9	04179DF375	369	1269,8	295,8	105	205,7
N10	0418CBEE7B	370	1882,8	332,1	175	232,1
Trung bình			1.746,9±272,0	359,9±53,1		210,1±33,9
Thông tin 10 mẫu cá thể tăng trưởng chậm						
STT	ID	Ngày tuổi	Khối lượng (g)	EBV	Gia đình	EBV gia đình
C1	0418CBDC66	384	767,2	-170,4	51	-120,9
C2	0418F7E641	363	606	-124,3	200	-58,1
C3	0418CC464D	349	230,4	-220,7	249	-68,4
C4	0418CBD251	398	516,6	-195	51	-120,9
C5	0418CBD623	393	657,8	-153,3	114	3,3
C6	041799FD22	403	500,2	-148,5	31	-78,9
C7	0418F7CC7B	371	394,2	-230,5	255	-171,6
C8	0416D9227A	378	447	-127,9	174	-39,9
C9	04179E0B12	405	445,6	-155,3	31	-78,9
C10	0418CBDA71	360	557,2	-110,4	296	-45,9
Trung bình			512,2±148,8	-163,6±40,6		-78,0±49,5

Tiếp theo, các SNP được sàng lọc từ bộ mẫu khởi tạo sẽ được xác định trên bộ mẫu kiểm nghiệm. Bộ mẫu kiểm nghiệm được thu thập bao gồm: 70 cá thể cá tra có chỉ số EBV cao nhất thuộc 24 gia đình có chỉ số EBV cao nhất theo thứ tự xếp hạng và 70 cá thể cá tra có chỉ số EBV thấp nhất thuộc 31 gia đình có chỉ số EBV thấp nhất theo thứ tự xếp hạng. Thông tin về khối lượng cơ thể và EBV của các cá thể này được trình bày trong Phụ lục 1.

Các mẫu vây của 160 cá thể này cũng được thu và bảo quản ngay trong cồn tuyệt đối ở nhiệt độ -20°C để sử dụng cho nghiên cứu.

2.2. Phương pháp

Luận án được tiến hành theo các bước được trình bày trong Hình 2.1.



Hình 2.1. Nội dung nghiên cứu và các nhóm phương pháp nghiên cứu tương ứng

2.2.1. Phân tích cấu trúc của các gen đích

Dựa trên trình tự các gen thuộc hệ thống IGF của cá tra nuôi đã được chú giải từ genome giải mã bằng NGS [16], các gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -

7 sẽ được phân tích cấu trúc để tìm các vùng quan trọng để xác định SNP. Trình tự protein suy diễn từ trình tự các gen đích được phân tích bằng ProteinBLAST (blastp) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>) với ngân hàng dữ liệu PDB, nhằm tìm ra các vùng bảo thủ quy định các domain đặc trưng. Ngoài ra, trình tự protein suy diễn cũng được phân tích bằng phần mềm SignalP5.0 [81] để dự đoán chuỗi peptit tín hiệu. Những vùng gen quy định các vùng domain chức năng, chuỗi peptit tín hiệu trên protein và các vùng gen liên quan đến điều hòa phiên mã, dịch mã sẽ được ưu tiên giải mã lại để xác định SNP.

2.2.2. Tách chiết DNA tổng số

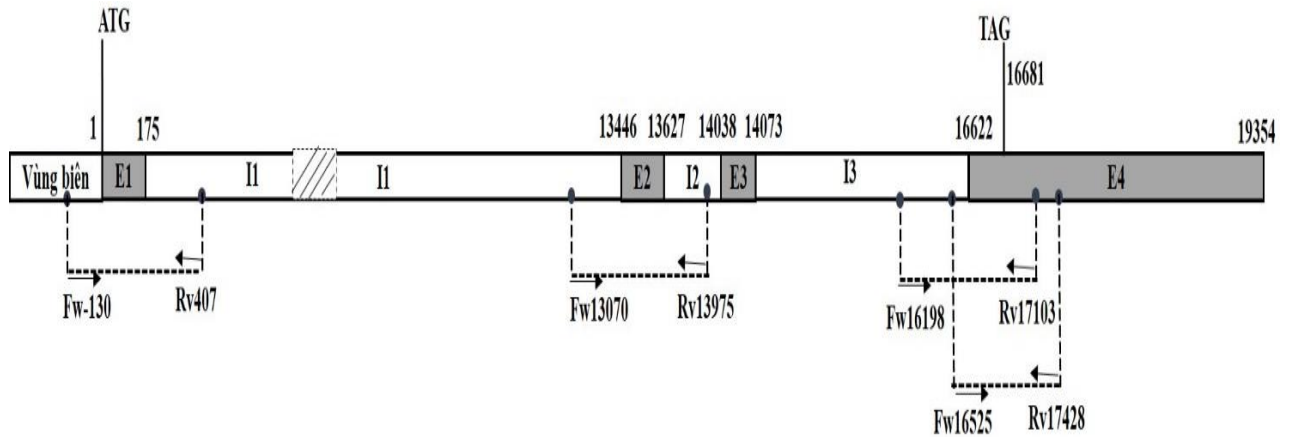
Các tế bào mô vây của cá tra được nghiền trong ni tơ lỏng thành dạng bột và đồng nhất trong dung dịch đệm (0,01 M EDTA, 0,01 M Tris-HCl pH 7,5, 0,1 M NaCl, 2,1% SDS và 100 µl/ml Proteinase K) ở 56°C trong 3 giờ. DNA tổng số được tách chiết từ dung dịch đồng nhất bằng phương pháp thường quy sử dụng phenol/chloroform [82]. Nồng độ và chất lượng DNA tổng số được kiểm tra 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

DNA tổng số tách chiết từ các mẫu vây cá tra được làm khuôn cho phản ứng PCR với các cặp mồi thiết kế để nhân các đoạn gen mong muốn. Năm mươi lăm cặp mồi được thiết kế bằng phần mềm Primer 3 (v.0.4.0) [83] để khuếch đại các đoạn gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 của cá tra nuôi dựa trên trình tự genome tham chiếu, với mã số định danh của các gen trên ngân hàng dữ liệu được trình bày trong Bảng 1.2. Vị trí của các đoạn gen được nhân lên và trình tự các cặp mồi được thiết kế cho các gen *IGF1*; *IGF2*; *IGF1R*; *IGFBP-1*; *IGFBP-2*; *IGFBP-3*, -5, -6 và *IGFBP-7* lần lượt được trình bày trong Hình 2.2/ Bảng 2.2, Hình 2.3/ Bảng 2.3, Hình 2.4/ Bảng 2.4, Hình 2.5/ Bảng 2.5, Hình 2.6/ Bảng 2.6, Hình 2.7/ Bảng 2.7, Bảng 2.8, Bảng 2.9 và Hình 2.8/ Bảng 2.10.

Phản ứng PCR được tiến hành để khuếch đại các đoạn gen với thể tích 25 µl gồm: 1 µl DNA khuôn, 1 µl mỗi mồi (10 pmol/µl) và 12,5 µl Taq 2X Master Mix (NEB). Điều kiện phản ứng như sau: biến tính chung ở 95°C 3 phút; 30 chu kỳ gồm bước biến tính ở

95°C 30 giây, điều kiện gắn mồi cho mỗi cặp mồi được trình bày trong các Bảng 2.2, 2.3, 2.4, 2.5, 2.6, 2.7, 2.8, 2.9, 2.10, kéo dài đoạn ở 68°C 1 phút; kéo dài đoạn chung 68°C 7 phút; bảo quản 4°C. Các sản phẩm được điện di kiểm tra kích thước trên gel agarose 1% và tinh sạch bằng kit GeneJET PCR Purification Kit (Thermo Fisher Scientific) theo hướng dẫn của nhà sản xuất.

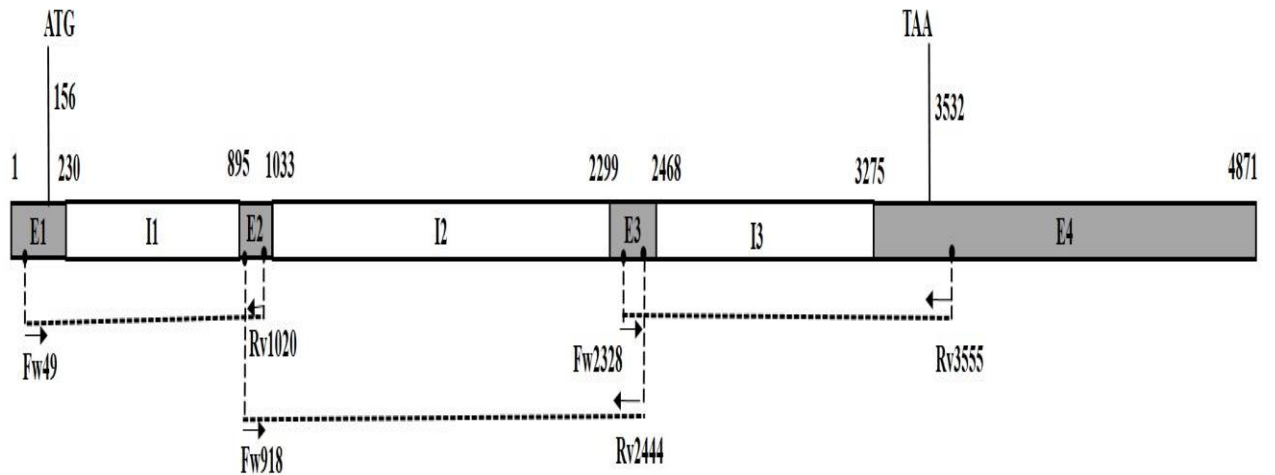


Hình 2.2. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGF1*

E1 đến E4: các exon, I1 đến I3: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.2. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGF1*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGF1</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
1	Fw-130	TGGAGTTAAATGTCTGCTTGTGA	537	53°C
	Rv407	TGGTATCAGTATCAGAGTCAAACG		30 giây
2	Fw13070	CGTTCTTAAATGTGTGATAAACTCC	906	51°C
	Rv13975	AAGTCACAAAAGCACTGATTTC		30 giây
3	Fw16198	GTGTCACTGCTTGTATTGTAAAGA	906	51°C
	Rv17103	AAAACCTTCACCTCTCCTCCA		30 giây
4	Fw16525	TTACACCACTCCGTCGTTGATTA	904	53°C
	Rv17428	TCCTGCTTAGCCATGTTCTCA		30 giây

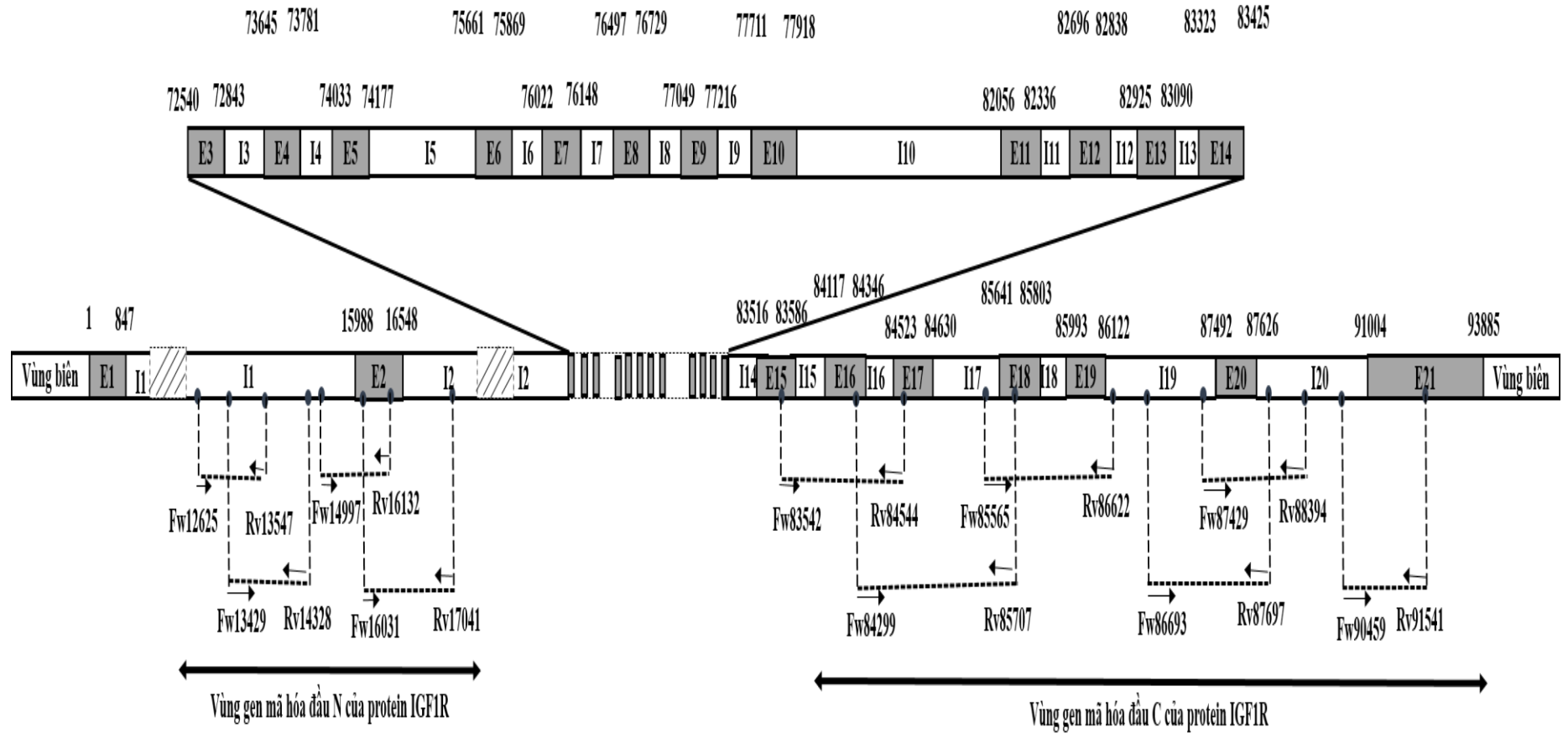


Hình 2.3. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGF2*

E1 đến E4: các exon, I1 đến I3: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.3. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGF2*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGF2</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gán mồi
5	Fw49	CCTCATCATCACCTCATCCTCATCC	972	62°C
	Rv1020	CTCTGTCTCCACACACGAACTGCAGAG		30 giây
6	Fw918	GCTTTTCACGGTGGCATTGTCACTG	1527	56°C
	Rv2444	GACGACCTGCAGCGTAGTGGTG		30 giây
7	Fw2328	CCTCACCGAGGAATAGTGGAGGAATGCTG	1228	56°C
	Rv3555	TAGGATCCCAAGGTCAAGGCTAATAATCC		30 giây

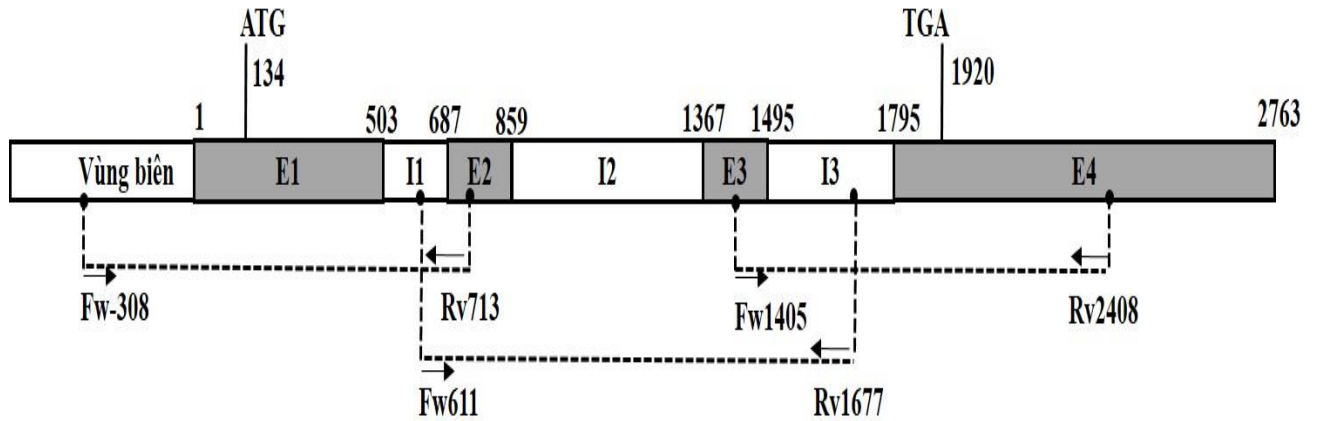


Hình 2.4. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGF1R*

E1 đến E21: các exon, I1 đến I20: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.4. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGF1R*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGF1R</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
8	Fw12625	TTATTTTCTGTAAGCCTGCACC	923	51°C
	Rv13547	GTGAATCCAACAGTGATCCGAG		30 giây
9	Fw13429	ACGGGGCTTTAGATGTATTTCTC	900	51°C
	Rv14328	TTCCAAACAGCCAAAACACTATC		30 giây
10	Fw14997	CTATACAGAGAACATGAGTCTCCAC	1136	53°C
	Rv16132	AAGCTAAGCATGCGAAACTCCTG		30 giây
11	Fw16031	CAAGAAGCTGGAGAACTGCACG	1011	54°C
	Rv17041	AGCAATATGACAGTTACTCCA		30 giây
12	Fw83542	TTCTCTATGCCTCTGTAAACCCA	1003	53°C
	Rv84544	AAACTACTCCCAGCAACCGTA		30 giây
13	Fw84299	TTCCTCAATGAAGCCTCTGTC	1409	56°C
	Rv85707	CGGAAATCTCGCCAGCCATC		15 giây
14	Fw85565	AGAGAAAGAAAAGGGAGTTGTAA	1058	51°C
	Rv86622	CTGGAGGATGTTAAGAGTGC		30 giây
15	Fw86693	GCTTTTGAAGTCCTCCAGATGAA	1005	51°C
	Rv87697	CAGTTTGAAGTGGTTATGTTTTGA		30 giây
16	Fw87429	ACATCTAATTGTGCTTTGCTCTCT	966	51°C
	Rv88394	CTGGATATGTGACGCCATCT		30 giây
17	Fw90459	GGCACTTTACTCCGTTTCAC	1083	53°C
	Rv91541	GACTCCATCTTTGCTCCTTGA		30 giây

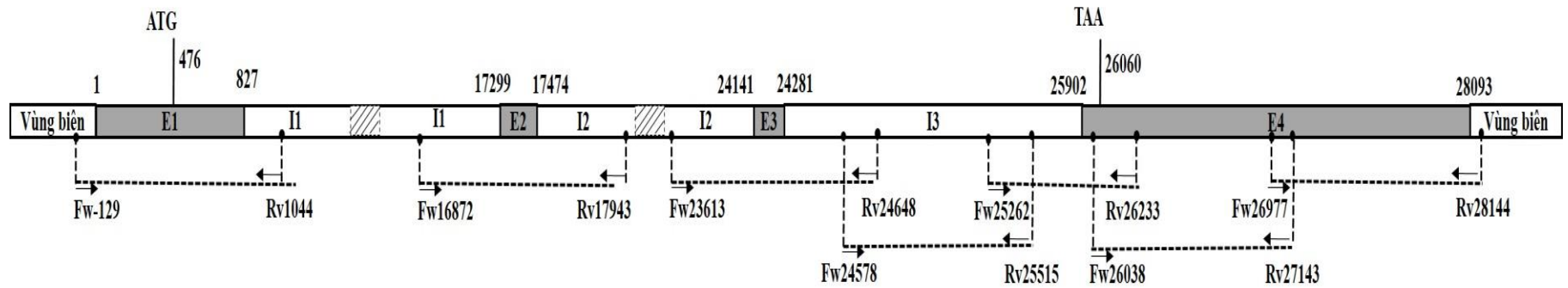


Hình 2.5. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGFBP-1*

E1 đến E4: các exon, I1 đến I3: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.5. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-1*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-1</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
18	Fw-308	GTTCACCCCTGCTGCTGTATG	1021	56°C
	Rv713	GACCATCACCTCCACCTC		30 giây
19	Fw611	CGCCGATTAGTGTCTGGAG	1067	51°C
	Rv1677	TACATCAGCCCAAACAATGTG		30 giây
20	Fw1405	GATTGCTCGATCCCAACAGAAG	1004	53°C
	Rv2408	GCAGTGACAATGTGGTTTGG		30 giây

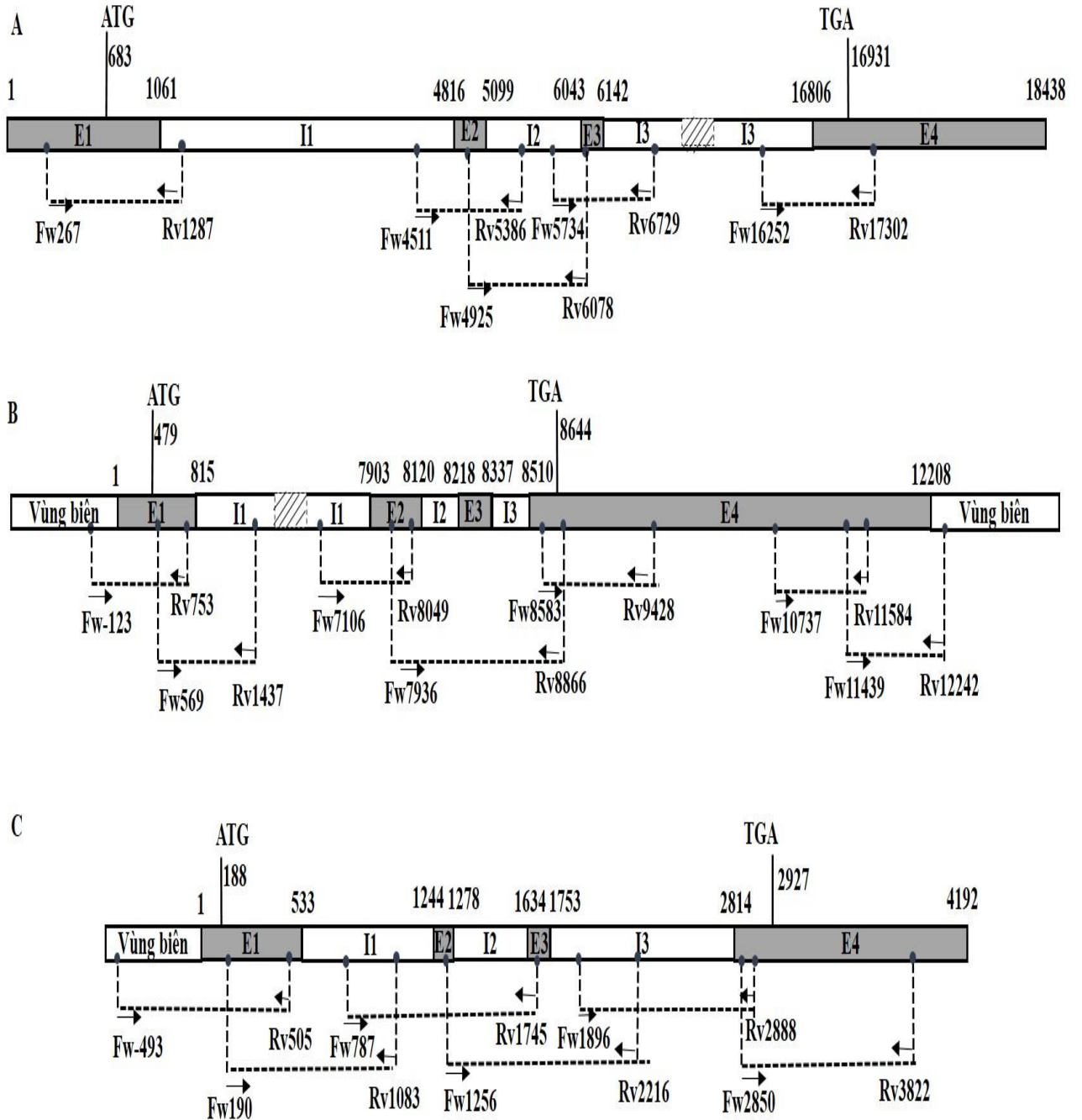


Hình 2.6. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGFBP-2*

E1 đến E4: các exon, I1 đến I3: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.6. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-2*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-2</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi	Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-2</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
21	Fw-129	GTATTTTTGGTGAACCCCTTG	1173	51°C 30 giây	25	Fw25262	GAGCATCCAAACATCCCAGTTC	972	53°C 30 giây
	Rv1044	TAGCCAAGATGTCAGCAAGC				Rv26233	CACCTGTTGTATGCTGGGTTTC		
22	Fw16872	TCCTGCTCAATGTTTGGTCA	1072	51°C 30 giây	26	Fw26038	CACCTTCAGCATCGCAAAC	1106	53°C 30 giây
	Rv17943	CAGTCAGCAATCCAGTTCATC				Rv27143	ACAGCTTCTAATCCAGCCA		
23	Fw23613	TGTTGGGAAAGGGTCAAATG	1036	51°C 30 giây	27	Fw26977	TACAGGATGTCCCAAAGGTC	1168	51°C 15 giây
	Rv24648	TCCTGCATATGGATCCTCAAC				Rv28144	CTGGCTGTCCTGAAAAGTGA		
24	Fw24578	TGTAAACCCAGGACAGTGA	938	53°C 30 giây					
	Rv25515	AAGGTGCAGGCTCTTTGTTG							



Hình 2.7. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGFBP-3* (A), *IGFBP-5* (B) và *IGFBP-6* (C).

E1 đến E4: các exon, I1 đến I3: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.7. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-3*

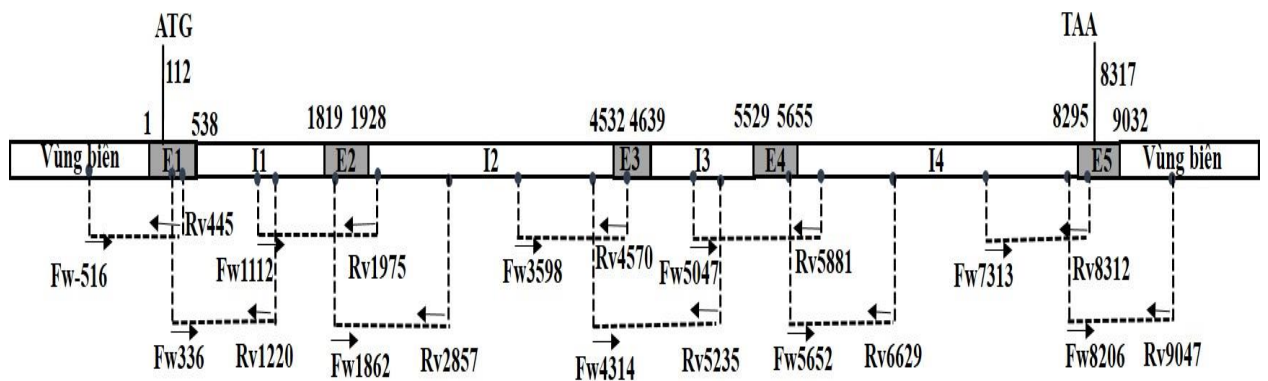
Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-3</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
28	Fw267	GGGCAGTTGTATAGCTCGTG	1021	54°C 30 giây
	Rv1287	TCCTGTAGAAGTTTCGTTTCGATT		
29	Fw4511	ACTGTTGAGTGTTGATCGGT	876	51°C 30 giây
	Rv5386	TCCTTCTCTTTCCTGACTCCTAATA		
30	Fw4925	CGGTATGCAGAAAACAGAGCTG	1154	51°C 30 giây
	Rv6078	CTTCAAAACGCTTTCATCTCTC		
31	Fw5734	AGAGAGGTGTGTTCAAGTGGTAT	996	53°C 30 giây
	Rv6729	TGAGTGAGAGCATTTACAAGCAG		
32	Fw16252	GCTTCTCCAAGTGCATATTTCT	1051	51°C 30 giây
	Rv17302	GCCAACCACATACTTTCAGTCA		

Bảng 2.8. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-5*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-5</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
33	Fw-123	CGTCATTTATTAGGGCGTCAGG	876	53°C 30 giây
	Rv753	AGTGCGTGTAGTGGCTTCTC		
34	Fw569	CAGAAGGCGCTGTCCATGTG	869	53°C 30 giây
	Rv1437	AGGAATGCAAGTGGGGATGT		
35	Fw7106	AAGTGTTTTTGGCCTGTTATTCCTC	944	53°C 30 giây
	Rv8049	GCTGCTTCTTCTTGTCCTTGCG		
36	Fw7936	AGACTGACACAACGGAGGAG	931	56°C 15 giây
	Rv8866	GGGACTCAGCTCGTAACACA		
37	Fw8583	CCGACTACAGTGGAGGGAAC	846	53°C 30 giây
	Rv9428	AGTGGGCCAGTGAGAGAATG		
38	Fw10737	AGCCATTCCCTGCCACAG	848	56°C 30 giây
	Rv11584	CCCCTTGACGGGGTAGTGA		
39	Fw11439	CACAGCTATTGATCTCAGTCCA	804	51°C 30 giây
	Rv12242	AACATTTTCACGGTCCTCCTG		

Bảng 2.9. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-6*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-6</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gán mồi
40	Fw-493	ATGCAAGAGTCAGATAGTTGAACG	998	53°C
	Rv505	CGTCCTGATCTGCTTGGTGC		30 giây
41	Fw190	GGAGTCCCACAGCGCAAAG	894	53°C
	Rv1083	GTGTATGAATCAGCACCATGAGAA		30 giây
42	Fw787	GCTGGGAATAGCTGTCAGTTTG	959	53°C
	Rv1745	TCCTGTAGTAGCCACGAGTGT		30 giây
43	Fw1256	CCCACAGCGGTGAAATTGAAA	961	51°C
	Rv2216	AGCAAAGCCTTGAAACTGAC		30 giây
44	Fw1896	AGTGGAGAAGTGTAGCTGAG	993	51°C
	Rv2888	TTGGGATGGTAGGGTGACG		30 giây
45	Fw2850	TGGTGCGTGGACGAATTAGG	973	56°C
	Rv3822	CCATCAAAGCCCGGTGTTCA		30 giây



Hình 2.8. Vị trí các cặp mồi được thiết kế trên trình tự gen *IGFBP-7*

E1 đến *E5*: các exon, *I1* đến *I4*: các intron. Vị trí của các cặp mồi xuôi (Fw) và mồi ngược (Rv) được đánh số theo vị trí trên gen.

Bảng 2.10. Danh sách các cặp mồi được thiết kế nhằm nhân các đoạn gen *IGFBP-7*

Ký hiệu cặp mồi	Mồi nhân gen <i>IGFBP-7</i>	Trình tự mồi (5'-3')	Kích thước sản phẩm dự kiến (bp)	Điều kiện gắn mồi
46	Fw-516	TCTGTGAGCCCCGACTTTCT	961	56°C
	Rv445	TGTACTCCACCCCGTCCGAG		30 giây
47	Fw336	TAAAGCCCTGGAGGACAAGA	885	51°C
	Rv1220	CTACGCTCGAAATCTGGACT		30 giây
48	Fw1112	GCTCTCTAACTCCTGCTTGGAC	864	53°C
	Rv1975	ACCTAGCGTGACATGGCTAA		30 giây
49	Fw1862	CGGTGCTCAGGTCTTCCTCA	996	53°C
	Rv2857	TCTGTTTTGGGTTTCCATGTTGTC		30 giây
50	Fw3598	TGACGAGCCAATGTGAAGAGG	973	56°C
	Rv4570	TCCGGGGAGCATCTGAGAG		30 giây
51	Fw4314	TCCTATAACAGCATGACAGCGA	922	53°C
	Rv5235	CATTTTCTGCGTGAGGGAACAG		30 giây
52	Fw5047	GTCTGTGTTGGTGCTGAGTTA	835	51°C
	Rv5881	GCTGCACATCATAGCACATTC		30 giây
53	Fw5652	AAGGGTGAGGTGCCAGGTTT	978	56°C
	Rv6629	CCTCTGACTGCCGTGATGC		30 giây
54	Fw7313	TCCCCCTCATAGCATTGACAG	1000	51°C
	Rv8312	GTCATCATCCTGGACCT		30 giây
55	Fw8206	TGCTGAACAACCAATCTGAG	842	51°C
	Rv9047	TGATCTTAAACCACGTCACCTC		30 giây

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

Các đoạn sản phẩm tinh sạch của các gen đích được dùng 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. Sản phẩm thu được được tinh sạch và rửa lại bằng EDTA và EtOH theo hướng dẫn của nhà sản xuất bộ kit trên, sau đó được bổ sung 10 μ l Hi-Di Formamide và biến tính ở 95°C trong 5 phút để điện di mao quản, xác định trình tự trên máy giải trình tự tự động ABI PRISM 3500 Avant Genetic Analyzer. Đối với mỗi đoạn gen, kết quả thô thu được từ máy giải trình tự sẽ được biên tập lại bằng phần mềm BioEdit [84] để thu được trình tự chính xác cuối cùng.

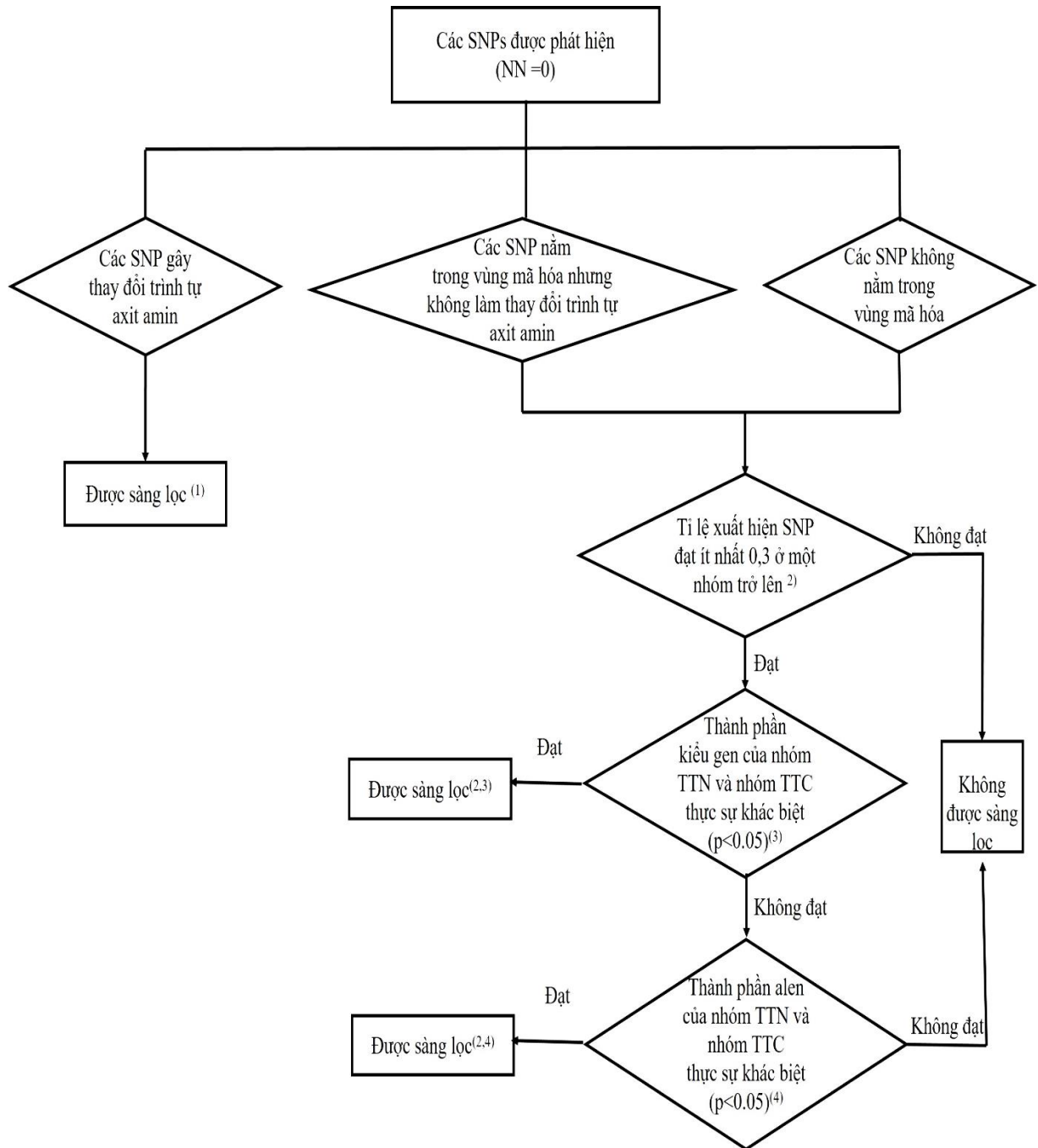
2.2.5. Kiểm tra tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham chiếu

Để xác định các đa hình, các trình tự tương ứng giải mã bằng Sanger của các cá thể khác nhau sẽ được so sánh với trình tự tham chiếu, do vậy, trước tiên, cần kiểm tra tính xác thực của kết quả giải mã bằng Sanger trên một cá thể so với trình tự tham chiếu. Theo đó, trình tự của các gen đích thuộc hệ thống IGF của một cá thể cá tra nuôi thu được giải mã lại bằng Sanger (thông qua các phương pháp ở mục 2.2.2, 2.2.3, 2.2.4) sẽ được so sánh với các trình tự tham chiếu thông qua NucleotideBLAST (BLASTn) (<https://blast.ncbi.nlm.nih.gov/Blast.cgi>), để tìm ra mức độ tương đồng. Mức độ tương đồng càng lớn chứng tỏ tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham chiếu càng cao, làm cơ sở để tiến hành nghiên cứu xác định đa hình trên gen đích theo hướng đã đề ra.

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

Trình tự nucleotide của các đoạn thuộc các gen đích tương ứng ở 10 cá thể cá tra tăng trưởng nhanh và 10 cá thể cá tra tăng trưởng chậm được giải mã bằng phương pháp Sanger (thông qua các phương pháp ở mục 2.2.1, 2.2.2, 2.2.3, 2.2.4) được so sánh (alignment) với trình tự tham chiếu tương ứng được giải mã bằng NGS [16] bằng phần mềm MUSCLE [85] để phát hiện các SNP.

SNP được xác định chính xác ở tất cả 20 cá thể thuộc bộ mẫu khởi tạo (số lượng các cá thể không đọc được trình tự ở vị trí SNP đó (Non-identify, ký hiệu NN) = 0 trong mỗi nhóm tăng trưởng nhanh/chậm) được quan tâm để đánh giá các tiêu chí sàng lọc.



Hình 2.9. Tiêu chí sàng lọc các SNP được phát hiện ở bộ mẫu khởi tạo

Với mục tiêu tìm kiếm SNP tiềm năng liên quan đến tính trạng tăng trưởng, từ những SNP được phát hiện và xác định chính xác ở tất cả 20 cá thể, luận án cần sàng lọc được những SNP có thể tác động đến chức năng của protein liên quan đến tính trạng, hoặc những SNP có tỉ lệ xuất hiện đủ lớn trong quần thể và đại diện cho nhóm tăng trưởng nhanh hoặc chậm thông qua sự khác biệt về thành phần kiểu gen và/hoặc thành phần alen giữa hai nhóm. Theo cơ sở này, tiêu chí sàng lọc SNP gồm: (1) gây nên sự

thay đổi axit amin trên protein tương ứng (Hình 2.9); hoặc (2) có tỉ lệ xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên và (3) thành phần kiểu gen giữa 2 nhóm khác biệt có ý nghĩa thống kê và/hoặc (4) thành phần alen giữa 2 nhóm khác biệt có ý nghĩa thống kê (giá trị $p < 0,05$ theo kiểm định Fisher Exact Test) (Hình 2.9). Cụ thể, tỉ lệ xuất hiện SNP ở mỗi nhóm tăng trưởng nhanh/chậm được xác định bằng tỉ số 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 (<https://www.well.ox.ac.uk/~wrayner/tools>). Alen thiểu số được xác định là alen có tần số nhỏ hơn trong tổng số alen của cả nhóm tăng trưởng nhanh và tăng trưởng chậm cộng lại. 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 software (<http://analysis.bio-x.cn/myAnalysis.php>) [86].

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

Để xác định 10 SNP được sàng lọc trên 140 cá thể sẽ cần 1400 phản ứng SBE đơn lẻ, đây là số lượng phản ứng lớn, cần nhiều thời gian, hóa chất, chi phí để tiến hành. Do đó, luận án chia 10 SNP được sàng lọc thành 2 nhóm để xác định thông qua phản ứng SBE multiplex (Bảng 2.11) để giảm số lượng phản ứng SBE xuống còn 2 phản ứng cho mỗi cá thể (tổng cộng 280 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.

DNA tổng số tách chiết từ mẫu vây của 70 cá thể cá tra tăng trưởng nhanh và 70 cá thể cá tra tăng trưởng chậm (theo phương pháp ở mục 2.2.2) được sử dụng làm khuôn cho phản ứng PCR (theo phương pháp ở mục 2.2.3, tuy nhiên thiết kế lại các cặp mồi) để nhân các đoạn gen chứa các SNP đã được sàng lọc (thu được từ phương pháp ở mục 2.2.6). Các cặp mồi cho phản ứng PCR được thiết kế lại bằng phần mềm Primer3 [83] để có thể thu được các đoạn gen chứa SNP được sàng lọc với kích thước ngắn và đồng đều nhau từ 500 đến 600 bp (Bảng 2.11).

Bảng 2.11. Trình tự môi nhân các đoạn chứa SNP được sàng lọc và môi SBE cho phản ứng kéo dài một nucleotide (SBE)

SNP được sàng lọc	Tên môi	Trình tự môi (5'-3')	Mục đích	Kích thước sản phẩm dự kiến
Nhóm 1				
<i>IGF1</i> . 13680 A>T	Fw13392	ATTCAGGGTCGTTCTGTACTC	Nhân đoạn chứa SNP	597 bp
	Rv13988	TCTGAAACCACTGAAGTCAC		
	SBE1	TCTCTAATCCAATATGTAAAAA GGAACATT	Phản ứng kéo dài 1 nucleotide	31 nucleotide
<i>IGF1R</i> . 13357 T>C	Fw12827	ATCGGTTTCTAGGCTGTTCC	Nhân đoạn chứa SNP	595 bp
	Rv13421	GCCCATTCACAAGGATTGC		
	SBE2	TTTTTTTTTTTTTTTTTTTTTTTTTTT TTTTTTATATCTTTGCATTTTAA TAATACCC	Phản ứng kéo dài 1 nucleotide	56 nucleotide
<i>IGF1R</i> . 15392 T>A	Fw15037	TTCCGACACATTGAGTTGC	Nhân đoạn chứa SNP	501 bp
	Rv15537	GGACGCCAATTCACCAGA		
	SBE3	TCTCTTCTCTTATTTTCAGACAG TATAGTCAAATAA	Phản ứng kéo dài 1 nucleotide	36 nucleotide
<i>IGF1R</i> . 83894 A>G	Fw83571	TTCAGTGCCGCAGAGAGTAA	Nhân đoạn chứa SNP	586 bp
	Rv84156	ATCTTCTCACGAGCCACCTC		
	SBE4	TTTTTTTTTTTTTTTTTCTCTACA AAATACAACCTTACACTATCTGC	Phản ứng kéo dài 1 nucleotide	46 nucleotide
Nhóm 2				
<i>IGFBP-3</i> . 704 C>G	Fw446	CTTGTTGCCCGTGCCTTTTC	Nhân đoạn chứa SNP	553 bp
	Rv998	GCAGCGGTTTGCTCTCAC		

	SBE5	GAACATGAAGCCCATATTCCG CTCT	Phản ứng kéo dài 1 nucleotide	26 nucleotide
<i>IGFBP-5.</i> 525 T>A	Fw204	GCTGGACCTTACTGGTGCTATC	Nhân đoạn chứa SNP	547 bp
	Rv750	GCGTGTAGTGGCTTCTCCTC		
	SBE6	TTTTTGTGATGGTGCCGTTTCT GTCGGCTG	Phản ứng kéo dài 1 nucleotide	31 nucleotide
<i>IGFBP-6.</i> 2278 C>A	Fw2199	CAGTTTCAAGGCTTTGCT	Nhân đoạn chứa SNP	627 bp
	Rv2825	AGAGGACCGACACTGAGGAA		
	SBE7	TCTCTCTCTCTCTCTCTCATT TTAAAGCCAAAAGTACAACA A	Phản ứng kéo dài 1 nucleotide	46 nucleotide
<i>IGFBP-7.</i> 344 T>C	Fw31	TCTGCGTCTCTACTCCGTC	Nhân đoạn chứa SNP	518 bp
	Fw548	CACTCCTCACCTCTGGCACA		
	SBE8	TTTTTTTTTTCTTCTTCAGCTTG TTCTTGTCCTCC	Phản ứng kéo dài 1 nucleotide	36 nucleotide
<i>IGFBP-7.</i> 2060 A>G	Fw1738	CACACACACCACATGACCTG	Nhân đoạn chứa SNP	551 bp
	Rv2288	GTTGGCAGAATGTCTGTTGG		
	SBE9	CTCTCTCTCTCTCAGGAGAT GTAGCCTAATATCTCAA	Phản ứng kéo dài 1 nucleotide	41 nucleotide
<i>IGFBP-7.</i> 4559 C>A	Fw4680	TCAACCCAGAGTGTAGCAG	Nhân đoạn chứa SNP	529 bp
	Rv4152	TGGGAAAATGGCTGTCAGTG		
	SBE10	TTTTTTTTTTTTTTTTTTTTTTT TCATGGACAATAAGAAGTTCT CTCAG	Phản ứng kéo dài 1 nucleotide	51 nucleotide

Các đoạn sản phẩm PCR chứa các SNP cần xác định thuộc cùng một nhóm (Bảng 2.11) của mỗi cá thể được dồn lại thành một mẫu, tinh sạch bằng kit GeneJET PCR Purification Kit (Thermo Fisher Scientific) theo hướng dẫn của nhà sản xuất để làm khuôn cho phản ứng SBE multiplex. Các môi SBE sử dụng cho phản ứng kiểm nghiệm

cũng được thiết kế chia làm 2 nhóm multiplex (Bảng 2.11), mỗi mỗi gồm phần trình tự lõi bắt cặp bổ sung với trình tự đích liền trước đầu 3' của SNP và phần trình tự đuôi là các đoạn lặp ở phía 5' của mỗi có chiều dài khác biệt 5 đến 8 nucleotide giữa các mỗi, nhằm phân biệt về chiều dài các sản phẩm thu được trong quá trình điện di mao quản tự động.

Phản ứng SBE được tiến hành sử dụng bộ kit ABI SNaPshot Multiplex PCR Kit (Applied BioSystems) theo hướng dẫn của nhà sản xuất. Sản phẩm SNaPshot được tinh sạch bằng 1 unit enzyme Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific) trong 37°C 30 phút, bất hoạt enzyme trong 65°C 15 phút. Sau đó, 0,5 µL sản phẩm tinh sạch được trộn cùng 9 µL Hi-Di Formamide và 0,5 µL GeneScan-120 LIZ Size Standard (Applied Biosystems) để biến tính ở 95°C 5 phút trước khi đưa sản phẩm vào hệ thống máy giải trình tự ABI 3500 Applied BioSystems để điện di mao quản. Dữ liệu thô được xử lý bằng phần mềm GeneMapper 5.1 (Applied BioSystems) sẽ cung cấp kết quả xác định SNP trên mỗi cá thể.

2.2.8. Phân tích dữ liệu

Dữ liệu được phân tích là bộ dữ liệu của 160 cá thể, gồm 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm (được tổng hợp từ dữ liệu xác định SNP của 140 cá thể thuộc bộ mẫu kiểm nghiệm và 20 cá thể thuộc bộ mẫu khởi tạo). Các phân tích hướng tới đánh giá đa dạng di truyền của 160 cá thể; đánh giá khả năng di truyền cùng nhau của các SNP được sàng lọc trên cùng một gen; 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 và nhóm tăng trưởng chậm; đánh giá tác động tổng hợp của các SNP lên tính trạng tăng trưởng thông qua sự khác biệt có ý nghĩa về tần số kiểu gen đơn bội haplotype tạo nên từ các SNP được sàng lọc trên cùng một gen hay từ các chỉ thị SNP tiềm năng giữa hai nhóm; và sự khác biệt có ý nghĩa về tần số tổ hợp kiểu gen tạo nên từ các SNP tiềm năng giữa hai nhóm. Các SNP có số lượng cá thể không đọc được trình tự SNP (Non-identify, ký hiệu NN) trong một nhóm tăng trưởng nhanh hoặc tăng trưởng chậm vượt quá 30% (NN>24 trong tổng số 80 cá thể của nhóm tăng trưởng nhanh/tăng trưởng chậm) sẽ bị loại.

Sự đa dạng di truyền của 160 cá thể bao gồm các thông tin đa hình (Polymorphism Information Content (PIC)), tần số alen thiểu số (Minor Allele Frequency (MAF)) được phân tích bằng phần mềm Gene-Calc (<https://www.gene-calc.pl>). Để có thể đại diện cho quần thể và có thể ứng dụng trong chọn giống, các SNP cần cho thấy mức độ đa dạng di truyền ở mức trung bình đến cao thông qua giá trị PIC nằm trong khoảng 0,25 đến 0,5 [87] và là các biến dị thường gặp với giá trị MAF trong khoảng từ 0,2 đến 0,5 [88].

Khả năng di truyền cùng nhau của các SNP trên cùng một gen được đánh giá thông qua sự mất cân bằng liên kết (linkage disequilibrium – LD). Sự mất cân bằng liên kết (LD) là sự kết hợp không ngẫu nhiên giữa các alen thuộc các locus khác nhau, được đánh giá dựa trên các giá trị D' và R^2 [89] với khoảng giá trị từ 0 đến 1. Giá trị D' hoặc R^2 bằng 0 cho thấy các alen ở hai locus khác nhau đạt trạng thái gần giống phân ly độc lập, còn khi D' hoặc R^2 càng gần đến 1 chứng tỏ chúng có khả năng cao được di truyền cùng nhau [89]. Các giá trị D' và R^2 được tính toán từ dữ liệu SNP trên 160 cá thể được tính toán bằng phần mềm SHEsis (<http://analysis.bio-x.cn/myAnalysis.php>) [86].

Sự khác biệt có ý nghĩa về các thông số hai nhóm tăng trưởng nhanh/chậm được xác định thông qua giá trị $p < 0,05$ theo kiểm định Fisher Exact Test. Cụ thể, sự khác biệt về thành phần kiểu gen, thành phần alen, tần số haplotype được phân tích bằng phần mềm SHEsis (<http://analysis.bio-x.cn/myAnalysis.php>) [86], với điều kiện mặc định ít nhất tần số của mỗi haplotype của một trong hai nhóm nhanh hoặc chậm phải lớn hơn 0,03. Sự khác biệt về tần số các tổ hợp kiểu gen tạo nên từ các SNP tiềm năng được phân tích bằng phần mềm NCSS 2021 Data Analysis (NCSS Statistic Software), với điều kiện mặc định ít nhất tỉ lệ của mỗi tổ hợp kiểu gen của một trong hai nhóm nhanh hoặc chậm phải lớn hơn 3%.

Trình tự gen *IGF1* phân tích được ở cá tra nuôi dài khoảng 19,4 kb chứa 4 vùng trình tự exon với chiều dài 175 bp, 182 bp, 36 bp, 2733 bp và 3 intron, trong đó cả 4 exon đều tham gia vào mã hóa cho các phần thuộc protein IGF1 (Hình 3.1 A, B). Theo kết quả BLAST với ngân hàng dữ liệu, trình tự này có mức độ tương đồng cao nhất với trình tự gen *IGF1* của cá nheo Mỹ *I. punctatus* ở mức độ 81,25%. Trong đó, một phần exon 1 quy định chuỗi peptit tín hiệu, một phần của exon 1 và một phần của exon 2 quy định domain IGF, một phần của exon 2, exon 3 và một phần nhỏ của exon 4 quy định domain E- là phần sẽ được loại bỏ khỏi tiền phối tử IGF để tạo thành IGF trưởng thành (Hình 3.1 A, B).

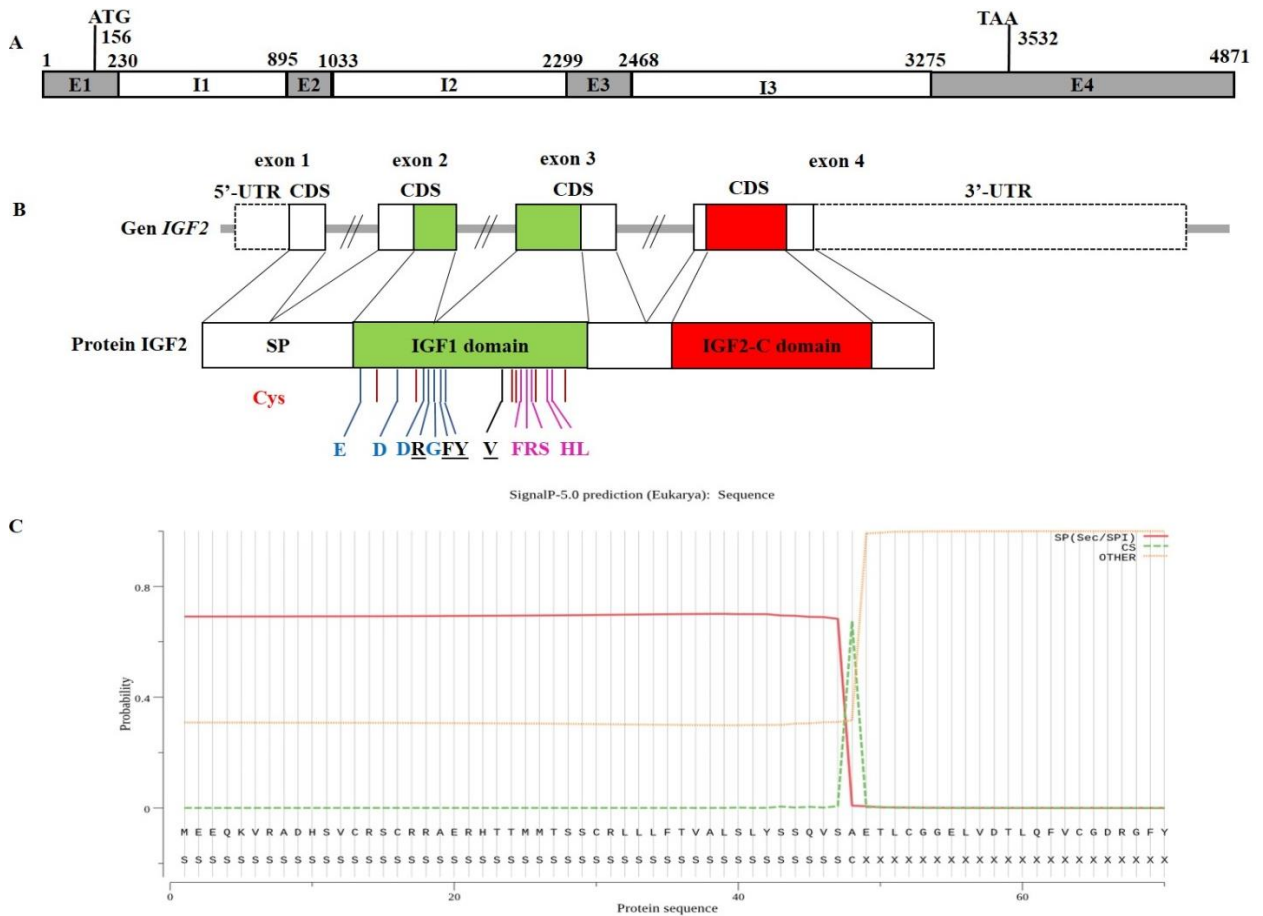
Ở protein, chuỗi peptit tín hiệu dài khoảng 33 axit amin, với điểm cắt của enzyme cắt chuỗi peptit tín hiệu nằm ở giữa axit amin thứ 33 và 34 là ARA-GP (theo dự đoán của phần mềm SignalP5.0 (Hình 3.1 C). Cấu trúc domain IGF của IGF1 trưởng thành chứa 6 phân tử cysteine nằm rải rác và cùng chứa các motif bám với IGFBP, thụ thể tuýp 1 và tuýp 2 của IGF/Insulin (Hình 3.1 B).

Toàn bộ exon 1 và exon 2 tham gia mã hóa cho chuỗi peptit tín hiệu và domain IGF1, cùng các vùng trình tự intron xung quanh hai exon này có thể liên quan đến điều hòa phiên mã, dịch mã sẽ được khuếch đại để tìm kiếm các SNP. Ngoài ra, vùng trình tự thuộc exon 4 có chứa codon kết thúc và gần 1000 nucleotide thuộc vùng 3'-UTR cùng intron liền kề trước exon này cũng được nhân lên vì có thể có chứa các trình tự điều hòa liên quan đến dịch mã của protein IGF1. Như vậy, có 4 cặp môi được thiết kế để khuếch đại các vùng quan trọng trên của gen *IGF1* để tìm kiếm các SNP (Hình 2.2, Bảng 2.2).

3.1.1.2. Gen và protein IGF2

Trình tự gen *IGF2* phân tích được ở cá tra nuôi chứa 4 vùng trình tự exon với chiều dài 230 bp, 139 bp, 170 bp và 1597 bp và 3 intron, trải dài 4,9 kb (Hình 3.2 A). Theo kết quả BLAST với ngân hàng dữ liệu, trình tự này có mức độ tương đồng cao nhất với trình tự gen *IGF2* của cá nheo Mỹ *I. punctatus* ở mức độ 94,22%. Trong đó cả 4 exon đều tham gia vào mã hóa cho các phần thuộc protein IGF2. Gen *IGF2* có một phần exon 1 và một phần exon 2 quy định chuỗi peptit tín hiệu, một phần của exon 2 và một

phần của exon 3 quy định domain IGF, còn exon 4 quy định domain IGF2-C (Hình 3.2 B).



Hình 3.2. Kết quả phân tích cấu trúc gen và protein IGF2 của cá tra nuôi

A: Cấu trúc gen gồm các exon và intron, với chiều dài tương ứng.

B: Cấu trúc gen tương ứng với các domain trên protein. SP: Signal peptide - chuỗi peptit tín hiệu. Cys: cysteine phân bố trong mỗi domain. Các đoạn thẳng đi kèm với trình tự axit amin: motif điển hình của domain.

C: Kết quả nhận biết cấu trúc chuỗi peptit tín hiệu của protein bằng phần mềm SignalP5.0.

Cấu trúc domain IGF của IGF2 giống với cấu trúc domain này ở IGF1 khi cùng chứa 6 phân tử cysteine nằm rải rác và cùng có chứa các motif bám với IGFBP, thụ thể týp 1 và týp 2 của IGF/Insulin, cho thấy sự tương đồng về chức năng của hai protein này. Chuỗi peptit tín hiệu dài khoảng 48 axit amin, với điểm cắt của enzyme cắt chuỗi peptit tín hiệu nằm ở giữa axit amin thứ 48 và 49 là VSA-ET (theo dự đoán của phần mềm SignalP5.0 (Hình 3.2 C)).

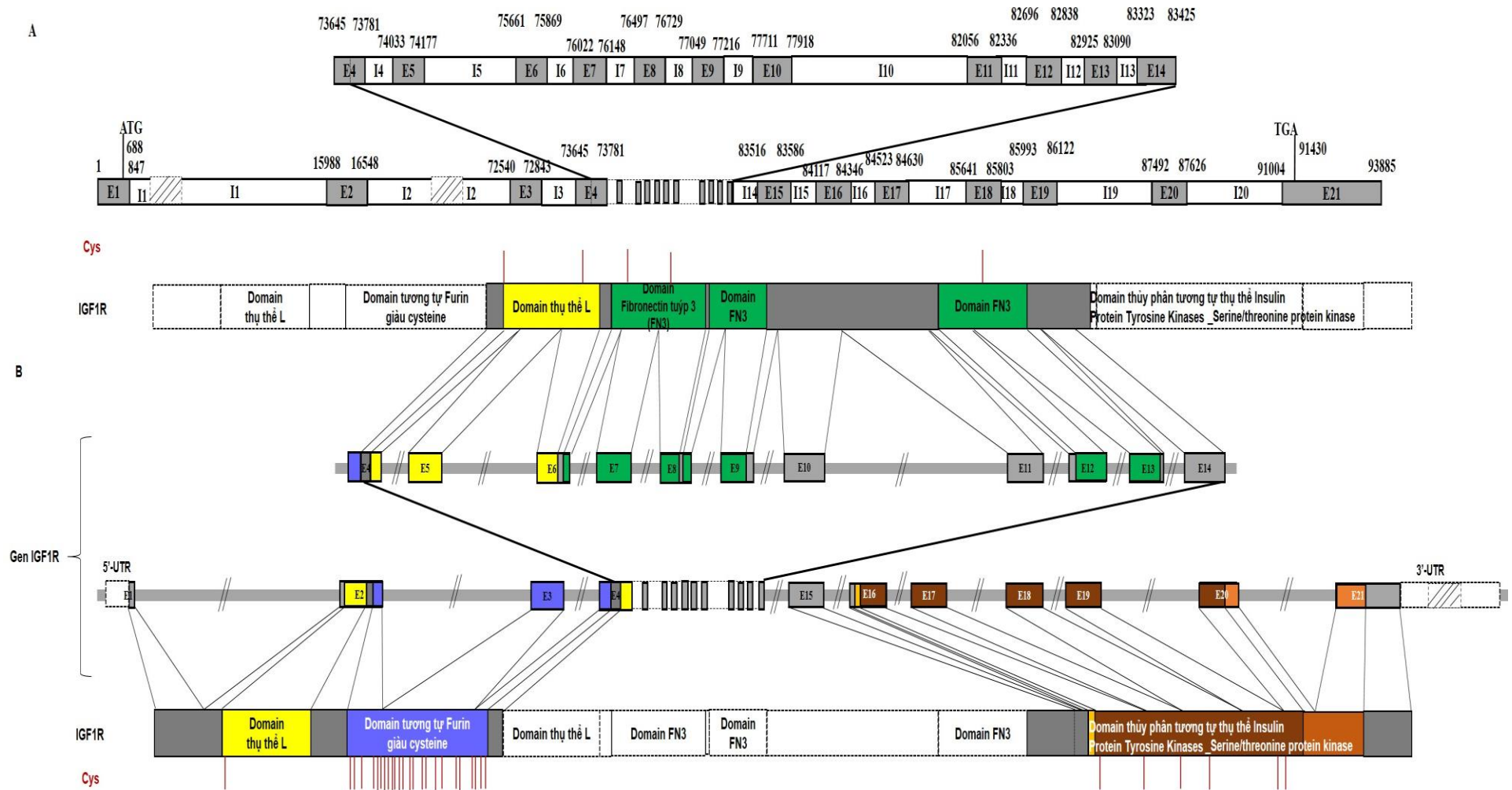
Do cả 4 exon đều tham gia mã hóa cho các vùng chức năng của protein, nên những vùng trình tự trên exon thuộc CDS, một phần nhỏ trình tự thuộc vùng 5-UTR, 3'-UTR và các intron xung quanh (có thể liên quan đến điều hòa phiên mã) của gen *IGF2* sẽ được khuếch đại bởi các cặp mồi được trình bày ở Hình 2.3, Bảng 2.3 để tìm kiếm SNP.

3.1.2. Gen mã hóa thụ thể *IGF1* (*IGF1R*) và protein tương ứng

Gen *IGF1R* của cá tra nuôi có 21 exon, 20 intron, trải dài 93,9 kb, với vùng mã hóa protein bắt đầu từ exon 1 và kết thúc ở exon 21 (Hình 3.3A). Chiều dài của các exon lần lượt là 848, 561, 304, 137, 145, 208, 127, 233, 168, 208, 281, 143, 166, 103, 71, 230, 108, 163, 130, 135 và 2882 bp. Khi tiến hành BLAST trình tự gen *IGF1R* của cá tra nuôi với ngân hàng dữ liệu, gen này được nhận thấy có mức độ tương đồng cao nhất đạt 91,37% với trình tự gen *IGF1R* của cá nheo Mỹ *I. punctatus*.

Trình tự protein IGF1R khi được BLAST với ngân hàng dữ liệu nhằm tìm các domain chức năng cho thấy 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, 3 domain Fibronectin tuýp 3 (FN3) và domain Tyrosine kinase. Theo đó, domain thụ thể L thứ nhất do exon 2 quy định, domain tương tự Furin giàu cysteine do một phần exon 2, exon 3 và một phần exon 4 quy định, domain thụ thể L thứ hai do một phần exon 4, exon 5 và một phần exon 6 quy định, domain FN3 thứ nhất do một phần exon 6, exon 7 và một phần exon 8 quy định, domain FN3 thứ hai do một phần exon 8 và exon 9 quy định, domain FN3 thứ ba do một phần exon 12 và exon 13 quy định và domain Tyrosine kinase do exon từ 16 đến 21 quy định (Hình 3.3 B).

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 [90], 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 sẽ được khuếch đại để tìm kiếm các SNP bằng các cặp mồi được trình bày ở Hình 2.4, Bảng 2.4.



Hình 3.3. Kết quả phân tích cấu trúc gen và protein IGF1R của cá tra nuôi

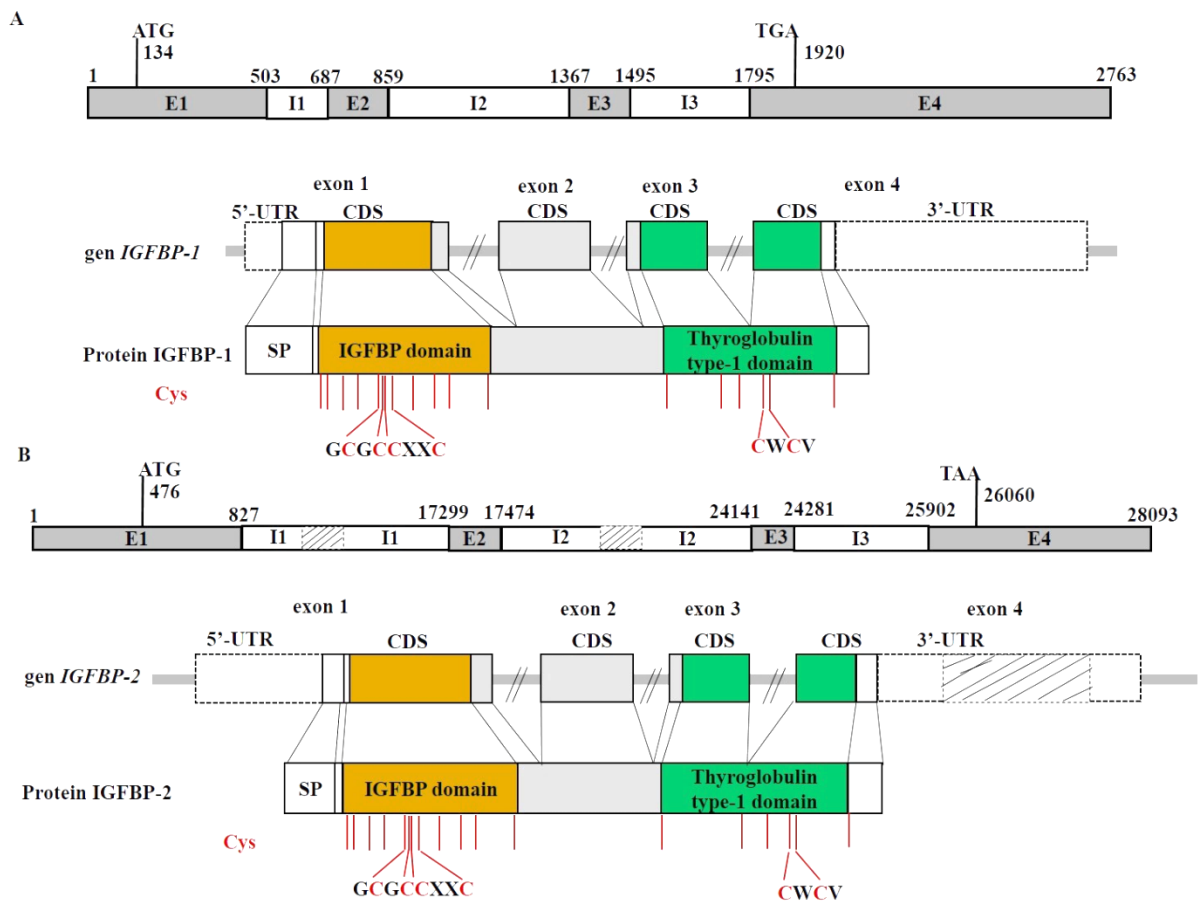
A: Cấu trúc gen gồm các exon và intron, với chiều dài tương ứng. B: Cấu trúc gen tương ứng với các domain trên protein.

3.1.3. Một số gen mã hóa IGFBP và các protein tương ứng

Cấu trúc các gen *IGFBP-1*, *-2*, *IGFBP-3*, *-5*, *-6* và *IGFBP-7* của cá tra nuôi được phân tích và biểu diễn trong các Hình 3.4, 3.5 và 3.6.

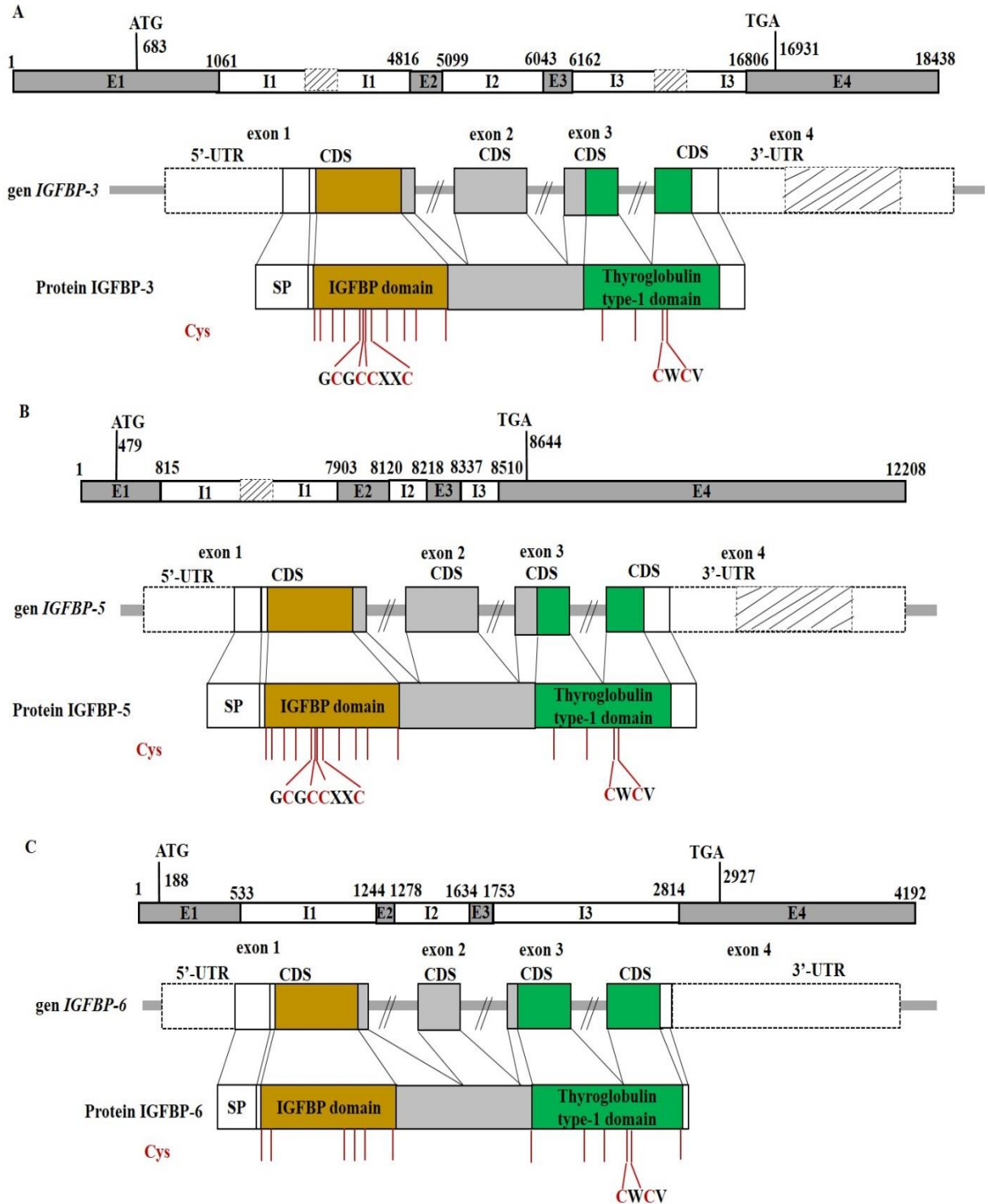
3.1.3.1. Gen và protein *IGFBP-1*, *-2*, *-3*, *-5*, *-6*

Các gen *IGFBP-1*, *-2*, *-3*, *-5*, *-6* đều chứa 4 exon và 3 intron, exon thứ nhất chứa codon khởi đầu và exon thứ tư chứa codon kết thúc (Hình 3.4, Hình 3.5). 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 với mức độ tương đồng về trình tự đạt cao nhất là 90,92%, 86,57%, 95,84%, 92,30% và 93,03% với các trình tự gen tương ứng của cá nheo Mỹ *I. punctatus*. Đó là kết quả thu được khi tiến hành BLAST trình tự các gen *IGFBP* của cá tra nuôi với ngân hàng dữ liệu./



Hình 3.4. Kết quả phân tích cấu trúc gen và protein IGFBP-1 (A) và IGFBP-2 (B).

Các đoạn trình tự axit amin chứa cysteine (C): các motif điển hình của mỗi domain.



Hình 3.5. Kết quả phân tích cấu trúc gen, protein IGFBP-3 (A), IGFBP-5 (B) và IGFBP-6 (C).

Các đoạn trình tự axit amin chứa cysteine (C): các motif điển hình của mỗi domain.

Kết quả phân tích cấu trúc các gen *IGFBP-1, -2, -3, -5, -6* được trình bày trong Hình 3.4, Hình 3.5 và Phụ lục 2 cho thấy các protein này đều chứa các vùng trình tự quy định IGFBP domain ở đầu N của protein, liền ngay sau đoạn peptit tín hiệu (SP) khoảng 3 đến 9 axit amin. Domain IGFBP này đều được quy định bởi trình tự nằm gọn trong exon 1 của mỗi gen tương ứng (Hình 3.4, 3.5) và đều chứa 12 cystein (C) bảo thủ tạo thành motif “GCGCCXXC” đặc trưng cho IGFBP (Hình 3.4, 3.5), trừ IGFBP6 chỉ chứa 6 cystein (Hình 3.5).

Domain thyroglobulin type 1 domain ở đầu C của các protein IGFBP-1, -2, -3, -5, -6 đều có chiều dài nhỏ hơn 80 axit amin và được mã hóa bởi vùng trình tự nằm trong exon 3 và 4 của các gen tương ứng. Domain này có chứa 5-6 cysteine bảo thủ, tạo thành motif “CWCV” (Hình 3.4, 3.5), góp phần tạo nên sự ổn định cấu hình không gian của các IGFBP trên.

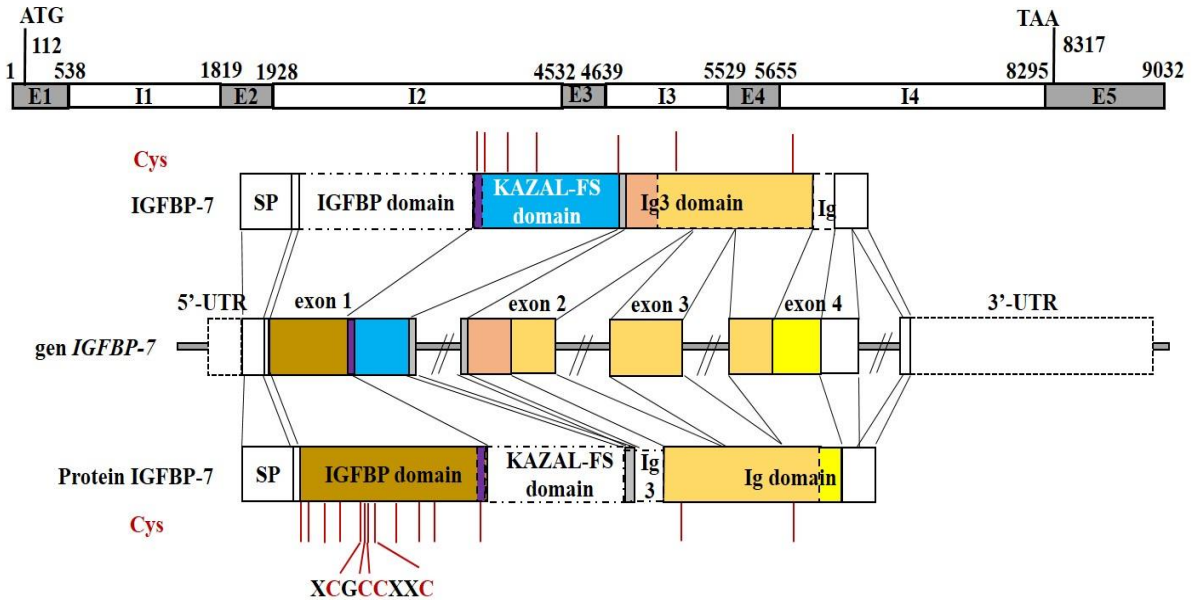
Như vậy, cấu trúc các gen và các protein IGFBP-1, -2, -3, -5, -6 ở cá tra nuôi rất tương đồng với nhau. Từ đó, các vùng gen thuộc CDS, vùng 5'-UTR, 3'-UTR cùng các vùng intron lân cận có thể liên quan đến điều hòa phiên mã của các gen này sẽ được lựa chọn để khuếch đại bằng các cặp môi được trình bày trong Hình 2.5/Bảng 2.5, Hình 2.6/Bảng 2.6 và Hình 2.7/Bảng 2.7, 2.8, 2.9.

3.1.3.2. Gen và protein IGFBP-7

Khác với các gen *IGFBP-1, -2, -3, -5, -6*, gen *IGFBP-7* có chiều dài khoảng 9 kb, chứa 5 exon và 4 intron, với exon thứ nhất chứa codon khởi đầu và exon thứ năm chứa codon kết thúc (Hình 3.6). Kết quả BLAST trình tự gen *IGFBP-7* của cá tra nuôi với ngân hàng dữ liệu cho thấy mức độ tương đồng cao nhất đạt 88,01% với các trình tự gen *IGFBP-7* của cá nheo Mỹ *I. punctatus*.

Đầu N của protein IGFBP-7 cũng chứa chuỗi peptit tín hiệu và domain IGFBP với motif điển hình tạo nên từ 12 cysteine bảo thủ như các protein IGFBP-1, -2, -3, -5, -6 (Hình 3.6). Domain IGFBP của protein này cũng được quy định bởi trình tự nằm gọn trong exon 1 của gen tương ứng. Tuy nhiên, khác với các protein IGFBP-1, -2, -3, -5, -6 có chứa domain nối ở giữa và domain thyroglobulin type 1 ở đầu C, protein IGFBP-7 lại chứa các domain KAZAL-FS, Ig3 và Ig, với các cysteine trong các domain ở đầu C phân bố rải rác, không tạo nên motif đặc trưng (Hình 3.6). Các domain này 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* sẽ được khuếch đại để tìm kiếm SNP, thông qua các cặp mồi được trình bày ở Hình 2.7/Bảng 2.10.



Hình 3.6. Kết quả phân tích cấu trúc gen và protein IGFBP-7.

Các đoạn trình tự axit amin chứa cysteine (C): các motif điển hình của domain.

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

Kết quả điện di DNA tổng số một cá thể được tách chiết (không trình bày cụ thể trong luận án) cho thấy DNA có chất lượng tốt với một băng điện di duy nhất, rõ nét, không bị đứt gãy, các chỉ số về nồng độ DNA, A260/A280 đủ tiêu chuẩn để sử dụng cho các thí nghiệm tiếp theo. Kết quả điện di sản phẩm PCR khuếch đại các đoạn gen bởi các cặp mồi và tinh sạch bằng phương pháp qua cột (không trình bày cụ thể trong luận án) cho thấy các băng thu được rõ nét, đặc hiệu với kích thước đúng như dự kiến, đủ tiêu chuẩn để sử dụng cho phản ứng giải trình tự bằng phương pháp Sanger.

Trình tự các đoạn gen sau khi được giải mã bằng phương pháp giải trình tự Sanger được so sánh với trình tự đoạn gen tham chiếu tương ứng thu được nhờ giải mã NGS [16] bằng phần mềm BLASTn cho thấy độ tương đồng về trình tự các đoạn gen tương ứng được giải mã bằng hai phương pháp nằm trong khoảng từ 99,9 đến 100 % (Phụ lục 3), khẳng định được tính xác thực của kết quả giải mã bằng Sanger so với trình tự tham

chiều. Do độ chính xác của hệ thống giải trình tự gen thế hệ mới có thể lên tới 99 % [91] và phương pháp giải mã bằng Sanger cũng được coi là tiêu chuẩn vàng trong số các phương pháp giải trình tự, nên kết quả này là hoàn toàn hợp lý, là 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

Hai mươi mẫu DNA tổng số được tách chiết từ 10 cá thể cá tra tăng trưởng nhanh và 10 mẫu cá tra tăng trưởng chậm thuộc bộ mẫu khởi tạo được điện di kiểm tra chất lượng trên gel agarose 1%. Kết quả (trình bày ở phần Phụ lục 4) cho thấy cả 20 mẫu đều có chất lượng tốt với một băng DNA tổng số duy nhất, rõ nét, không bị đứt gãy, đủ tiêu chuẩn để sử dụng cho các thí nghiệm tiếp theo.

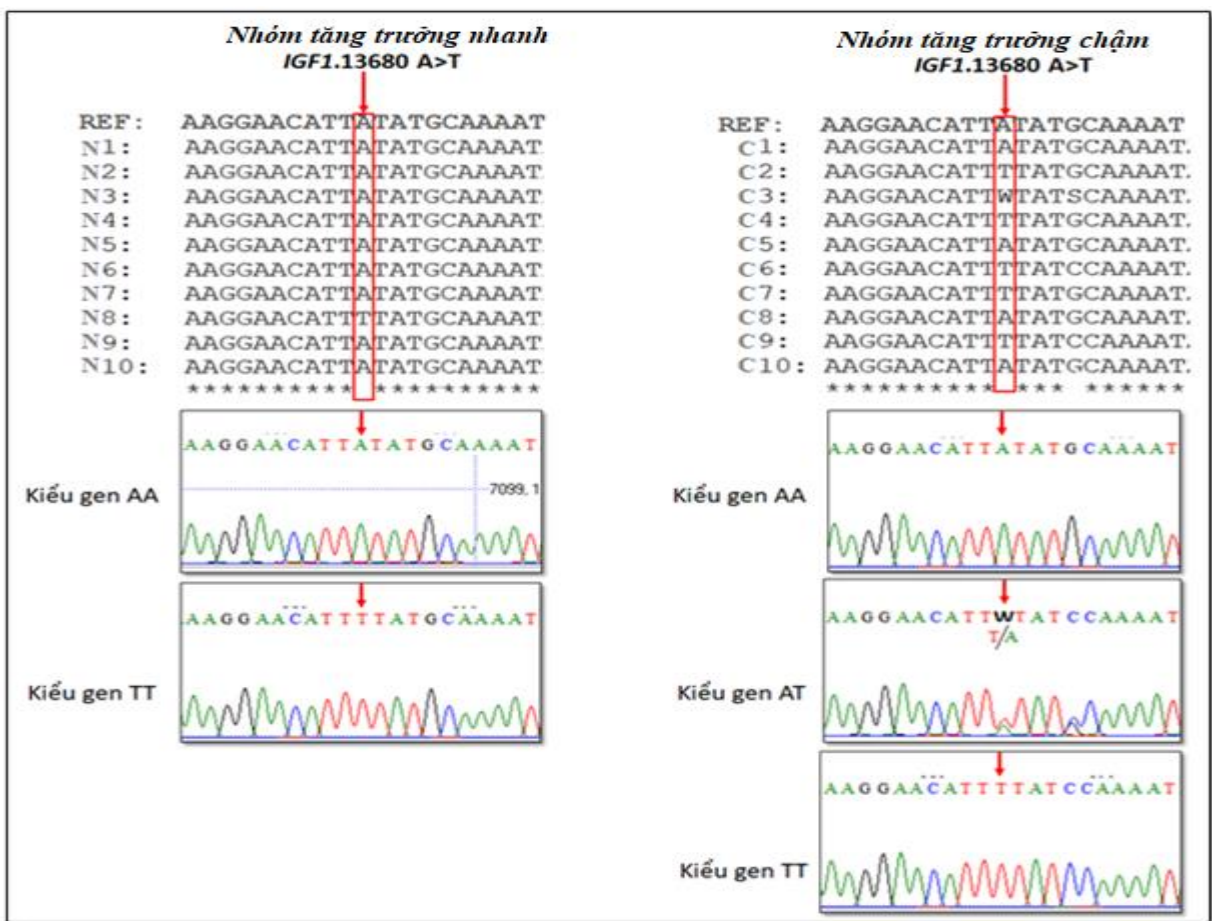
Kết quả điện di sản phẩm PCR khuếch đại từ 55 cặp môi nhân các đoạn gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 được trình bày trong Phụ lục 5, cho thấy các băng thu được rõ nét, đặc hiệu với kích thước đúng như dự kiến, đủ tiêu chuẩn để sử dụng cho phản ứng giải trình tự bằng phương pháp Sanger.

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

Trình tự các đoạn gen *IGF1* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu để xác định SNP. Định tín hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được minh họa ở Hình 3.7. Có 10 SNP được phát hiện trên gen *IGF1*, tất cả đều nằm trên intron (Bảng 3.1).

Theo tiêu chí sàng lọc, 10 SNP nằm trong vùng không mã hóa này trước tiên sẽ được kiểm tra tỉ lệ xuất hiện SNP ở hai nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm. Kết quả cho thấy có 4 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 tỉ lệ xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên (Bảng 3.1). Do vậy, 4 SNP này được tiếp tục đánh giá sự khác biệt về thành phần kiểu gen/thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm thông qua giá trị p nhỏ hơn 0,05. Kết quả, chỉ có SNP 13680 A>T thể hiện sự khác biệt về thành phần kiểu gen giữa nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm, với giá trị p là 0,03.

Cụ thể, nhóm tăng trưởng nhanh có 2 kiểu gen phát sinh từ SNP này là AA và TT, với kiểu gen AA chiếm ưu thế, trong khi nhóm tăng trưởng chậm lại có 3 kiểu gen là AA, AT và TT (Hình 3.7), với TT chiếm tỉ lệ cao nhất (Bảng 3.1). Thành phần alen của SNP này ở hai nhóm tăng trưởng nhanh/ chậm cũng có sự khác biệt có ý nghĩa, với giá trị p là 0,005. Ở nhóm tăng trưởng nhanh, alen A chiếm tỉ lệ cao hơn rõ rệt (18/20 alen), còn ở nhóm tăng trưởng chậm, alen T lại có tỉ lệ cao hơn (11/20 alen) (Bảng 3.1). Như vậy, SNP *IGF1*.13680 A>T sẽ được sàng lọc để xác định trên bộ mẫu kiểm nghiệm nhằm phân tích sự liên quan của nó đến tính trạng tăng trưởng của cá tra nuôi.



Hình 3.7. Kết quả phát hiện *IGF1*.13680 A>T trên gen *IGF1* 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.

Bảng 3.1. Danh sách 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; (^a): Tỷ lệ xuất hiện SNP của mỗi nhóm, ⁽²⁾, ⁽³⁾, ⁽⁴⁾: đáp ứng tiêu chí sàng lọc SNP số 2, 3, 4 tương ứng.

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

Trình tự các đoạn gen *IGF2* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu, từ đó phát hiện được 12 SNP, trong đó có 1 SNP thuộc vùng mã hóa nhưng không làm thay đổi trình tự axit amin trên protein và 11 SNP thuộc các intron (Bảng 3.2).

Bảng 3.2. Danh sách các SNP được phát hiện trên gen *IGF2*

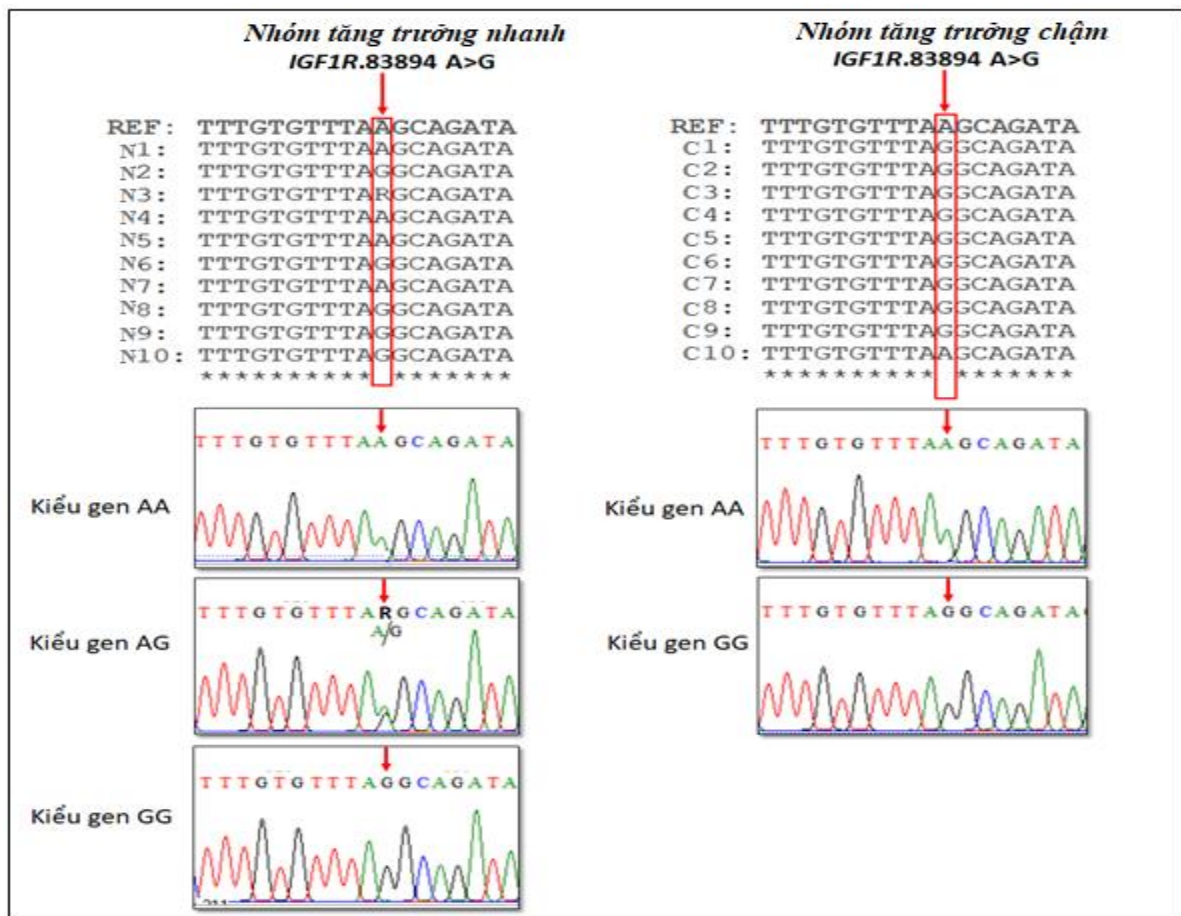
ST T	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	E1_ CDS_ 179	A	G*	9AA:1NN	7AA:2GA:1NN	18A	16A:2G
2	I1_237	G	A*	8GG:1GA:1NN	5GG:4GA:1NN	17G:1A	14G:4A
3	I2_ 1355 ⁽²⁾	A*	G	3AA:4AG:3GG (1,00)	2AA:2AG:6GG (0,5)	10A:10G	6A:14G
				p = 0,45		p = 0,33	
4	I2_ 1391 ⁽²⁾	G*	A	4GG:3GA:3AA (1,16)	3GG:2GA:5AA (0,71)	11G:9A	8G:12A
				p = 0,63		p = 0,53	
5	I2_1393	T	A*	8TT:2AA (0,25)	10TT (0,00)	16T:4A	20T
6	I2_1674	G*	C	1GC:6CC:3NN	1GG: 9CC	1G:13C	2G:18C
7	I2_1736	A	T*	6AA:1AT:3NN	8AA:2NN	13A:1T	16A
8	I3_ 2601 ⁽²⁾	T	A*	8TT:2TA (0,20)	7TT: 3AT (0,30)	18T:2A	17T:3A
				p = 1,0		p = 1,0	
9	I3_2685	A	T*	10AA (0,00)	8AA: 2AT (0,20)	20A	18A:2T
10	I3_2794	G	A*	10GG (0,00)	8GG:2GA (0,20)	20G	18G:2A
11	I3_2819	A	T*	10AA (0,00)	8AA:2TA (0,20)	20A	18A:2T
12	I3_3061	C	A*	10CC (0,00)	8CC:2CA (0,20)	20C	18C:2A

*: alen thiếu số; (^a): Tỷ lệ xuất hiện SNP của mỗi nhóm, (²): đáp ứng tiêu chí sàng lọc SNP số 2.

Theo tiêu chí sàng lọc SNP, có 8 trên 12 SNP được xác định trên toàn bộ 20 cá thể (NN =0) sẽ được đánh giá tỷ lệ xuất hiện SNP ở hai nhóm tăng trưởng nhanh/chậm cần đạt ít nhất 0,3 ở một nhóm trở lên. Kết quả cho thấy có 3 SNP đạt tiêu chí này là 1355 A>G, 1391 G>A ở intron 2 và 2061 T>A ở intron 3 (Bảng 3.2). Tuy nhiên, 3 SNP này không thể hiện được sự khác biệt có ý nghĩa về thành phần kiểu gen và/hoặc thành phần alen giữa hai nhóm thông qua giá trị p nhỏ hơn 0,05 (Bảng 3.2). Do đó không có SNP nào trên gen *IGF2* được sàng lọc để tiếp tục được xác định ở bộ mẫu kiểm nghiệm.

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

Trình tự các đoạn gen *IGF1R* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu để xác định SNP. Định tin hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được minh họa ở Hình 3.8. Tất cả 9 SNP được phát hiện ở gen *IGF1R* hoàn toàn nằm trong các intron (Bảng 3.3). Chín SNP này đều có tỉ lệ xuất hiện SNP đạt ít nhất 0,3 ở một nhóm tăng trưởng nhanh/chậm trở lên (Bảng 3.3), do đó tất cả các SNP tiếp tục được đánh giá sự khác biệt về thành phần kiểu gen/ thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm thông qua giá trị p nhỏ hơn 0,05.



Hình 3.8. Kết quả phát hiện *IGF1R.83894 A>G* trên gen *IGF1R* 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.

Bảng 3.3. Danh sách 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	I1_12992 ⁽²⁾	A	C*	9AA:1AC (0,10)	6AA:2AC:2CC (0,50)	19A:1C	14A:6C
				p = 0,29		p = 0,09	
2	I1_13357 ^(2,3)	T	C*	4TT:6TC (0,60)	8TC:2CC (1,25)	14T:6C	8T:12C
				p = 0,04		p = 0,11	
3	I1_15359 ⁽²⁾	C	G*	7CC:3GG (0,43)	4CC:6GG (1,50)	14C:6G	8C:12G
				p = 0,37		p = 0,11	
4	I1_15392 ^(2,3,4)	T	A*	5TT:5TA (0,50)	10TT (0,00)	15T:5A	20T
				p = 0,03		p = 0,04	
5	I1_15397 ⁽²⁾	T	G*	5TT:3TG:2GG (0,63)	8TT:2TG (0,20)	13T:7G	18T:2G
				p = 0,23		p = 0,12	
6	I15_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	I17_85057 ⁽²⁾	T	C*	10TT (0,00)	7TT:1TC:2CC (0,38)	20T	16T:4C
				p = 0,11		p = 0,1	
8	I19_87147 ⁽²⁾	T	C*	10TT (0,00)	7TT:3TC (0,30)	20T	17T:3C
				p = 0,21		p = 0,23	
9	I20_90590 ⁽²⁾	A	T*	7AA:3AT (0,30)	10AA (0,00)	17A:3T	20A
				p = 0,21		p = 0,23	

*: alen thiếu số; (^a): Tỷ lệ xuất hiện SNP của mỗi nhóm, (²), (³), (⁴): đáp ứng tiêu chí sàng lọc SNP số 2, 3, 4 tương ứng.

Theo đó, có 3 SNP thể hiện được 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 (Bảng 3.3).

Cụ thể, đối với SNP 13357T>C ở intron 1, nhóm tăng trưởng nhanh có kiểu gen TT và TC, trong khi nhóm tăng trưởng chậm có kiểu gen TC và CC, với giá trị p là 0,04 (Bảng 3.3). SNP 15392 T>A ở intron 1 thể hiện cả sự khác biệt về thành phần kiểu gen và thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, với giá trị p lần lượt là 0,03 và 0,04. Với SNP này, có 2 kiểu gen trong nhóm tăng trưởng nhanh là TT và TA với tỉ lệ ngang nhau, trong khi chỉ có 1 kiểu gen TT trong nhóm sinh chậm, dẫn tới tỉ lệ alen T và A ở hai nhóm cũng có sự khác biệt (Bảng 3.3). Ở SNP 83894 A>G trên intron 15, thành phần alen giữa hai nhóm có sự khác biệt với p bằng 0,03, cụ thể tỉ lệ alen A/G (alen thiểu số/alen đa số) của nhóm tăng trưởng nhanh (0,81) cao hơn hẳn nhóm tăng trưởng chậm (0,11) (Bảng 3.3). Như vậy, 3 SNP 13357 T>C, 15392T>A trên intron 1 và 83894 A>G trên intron 15 của gen *IGF1R* được sàng lọc để xác định trên bộ mẫu kiểm nghiệm.

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

Trình tự các đoạn gen *IGFBP-1* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu, từ đó phát hiện được 4 SNP, trong đó có 1 SNP thuộc vùng promoter và 3 SNP thuộc các vùng intron của gen (Bảng 3.4).

Bảng 3.4. Danh sách các SNP được phát hiện trên gen *IGFBP-1*

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	Promoter_ -35 ⁽²⁾	G*	C	1GG:4GC:5CC (0,56)	2GG:6GC:2CC (0,50)	6G:14C	10G:10C
				p = 0,40		p = 0,33	
2	I2_1053 ⁽²⁾	A	T*	7AA:3AT (0,30)	7AA:2AT:1TT (0,33)	17A:3T	16A:4T
				p = 0,81		p = 0,99	
3	I2_1333 ⁽²⁾	T	C*	5TT:4TC:1CC (0,56)	5TT:4TC:1CC (0,56)	14T:6C	14T:6C
				p = 1,0		p = 1,0	
4	I3_1659 ⁽²⁾	C	G*	5CC:4CG:1GG (0,56)	4CC:5CG:1GG (0,67)	14C:6G	13C:7G
				p = 1,0		p = 0,99	

*: alen thiểu số; (^a): Tỉ lệ xuất hiện SNP của mỗi nhóm, (²): đáp ứng tiêu chí sàng lọc SNP số 2.

Theo tiêu chí sàng lọc SNP, 4 SNP nằm trong vùng không mã hóa trên sẽ được đánh giá tỉ lệ xuất hiện SNP cần đạt ít nhất 0,3 ở một nhóm trở lên và kết quả cho thấy cả 4 SNP đều đạt tiêu chí này (Bảng 3.4). Tuy nhiên, không có SNP nào có sự khác biệt có ý nghĩa về thành phần kiểu gen và/hoặc thành phần alen giữa hai nhóm thông qua giá trị p nhỏ hơn 0,05 (Bảng 3.4). Do đó, không có SNP nào trên gen *IGFBP-1* được sàng lọc để xác định ở bộ mẫu kiểm nghiệm.

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

Trình tự các đoạn gen *IGFBP-2* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu, từ đó phát hiện được 10 SNP, trong đó có 1 SNP thuộc vùng exon với 1 SNP 787 G>A nằm trong vùng mã hóa nhưng không làm thay đổi trình tự axit amin và 2 SNP nằm trong vùng 3'-UTR và 7 SNP thuộc các vùng intron của gen (Bảng 3.5).

Bảng 3.5. Danh sách các SNP được phát hiện trên gen *IGFBP-2*

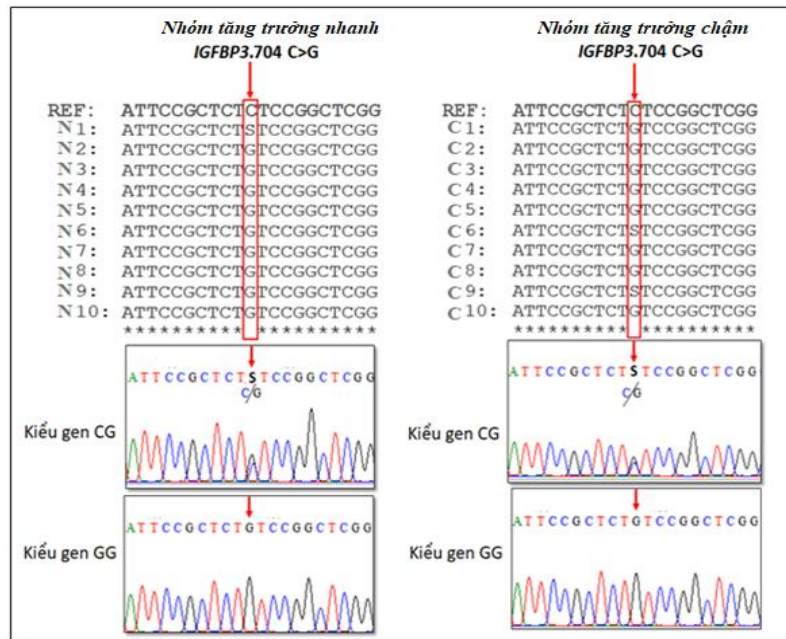
STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	E1_CDS_787	G*	A	3GA:7AA	3GA:6AA:1NN	3G:17A	3G:15A
2	I2_23777	C	T*	10CC	7CC:1CT:2NN	20C	15C:1T
3	I2_23916	G*	A	4GA:6AA	3GA:5AA:2NN	4G:16A	3G:13A
4	I3_24467	T	A*	7TT:3TA	8TT:2NN	17T:3A	16T
5	I3_24480	T	G*	10TT	7TT:1GG:2NN	20T	14T:2G
6	I3_24542	T	G*	10TT	7TT:1GG:2NN	20T	14T:2G
7	I3_25376	A	T*	6AA:3AT:1NN	8AA:1TT:1NN	15A:3T	16A:2T
8	I3_25443	G	A*	9GG:1NN	6GG:2GA:1AA:1NN	18G	14G:4A
9	E4_3'UTR 26814 ⁽²⁾	A	T*	8AA:2AT (0,20)	9AA:1AT (0,10)	18A:2T	19A:1T
10	E4_3'UTR 27414 ⁽²⁾	C	T*	6CC:4CT (0,40)	8CC:2CT (0,20)	16C:4T	18C:2T
				p = 0,63		p = 0,66	

*: alen thiểu số; (^a): Tỉ lệ xuất hiện SNP của mỗi nhóm, (²): đáp ứng tiêu chí sàng lọc SNP số 2.

Các SNP chỉ được đưa vào sàng lọc khi được xác định chính xác trên toàn bộ 20 cá thể. Do đó, trong 10 SNP được phát hiện, chỉ có 2 SNP 26814 A>T và 27414 C>T thuộc vùng 3'-UTR ở exon 4 được tiếp tục đánh giá tỉ lệ xuất hiện SNP ở nhóm tăng trưởng nhanh/chậm. Kết quả cho thấy chỉ có SNP 27414 C>T có tỉ lệ xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên (Bảng 3.5) và được đánh giá sự khác biệt về thành phần kiểu gen/thành phần alen giữa hai nhóm thông qua giá trị p nhỏ hơn 0,05. Tuy nhiên, các giá trị p đều không đạt nhỏ hơn 0,05 (Bảng 3.5), cho thấy không có sự khác biệt về thành phần kiểu gen và /hoặc thành phần alen phát sinh từ SNP này giữa hai nhóm. Do đó không có SNP nào trên gen *IGFBP-2* đạt đủ tiêu chí để được sàng lọc.

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

Trình tự các đoạn gen *IGFBP-3* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu để xác định SNP. Định tin hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được phát hiện được minh họa ở Hình 3.9.



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.

Trong các vùng được khảo sát của gen, 10 SNP được phát hiện gồm 6 SNP thuộc các intron và 4 SNP thuộc các exon. Các SNP thuộc exon nằm trong các vùng 5'-UTR, CDS và 3'-UTR (Bảng 3.6).

Bảng 3.6. Danh sách các SNP được phát hiện trên gen *IGFBP-3*

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	E1_5'UTR_647	T	A*	8TT:2AA (0,25)	8TT:1TA:1AA (0,22)	16T:4A	17T:3A
2	E1_CDS_704 ⁽¹⁾	C*	G	1CG:9GG	2CG:8GG	1C:19G	2C:18G
3	I1_1115	G	T*	9GG:1TT (0,11)	9GG:1TT (0,11)	18G:2T	18G:2T
4	I1_1123 ⁽²⁾	G*	T	2GG:8TT (0,20)	3GG:7TT (0,43)	4G:16T	6G:14T
				p = 1,0		p = 0,72	
5	I1_4638	C	G*	8CC:2CG (0,20)	10CC (0,00)	18C:2G	20C
6	I1_4695	C	A*	9CC:1CA (0,01)	8CC:2CA (0,20)	19C:1A	18C:2A
7	I2_5187	G	A*	10GG (0,00)	9GG:1GA (0,10)	20G	19G:1A
8	I2_5375	G	A*	6GG:2GA:2NN	8GG:1GA:1NN	14G:4A	17G:1A
9	E4_3'UTR_17119 ⁽²⁾	A*	G	3AA:7GG (0,43)	2AA:8GG (0,25)	6A:14G	4A:16G
				p = 1,0		p = 0,72	
10	E4_3'UTR_17128	C	G*	8CC:2GG (0,25)	10CC (0,00)	16C:4G	20C

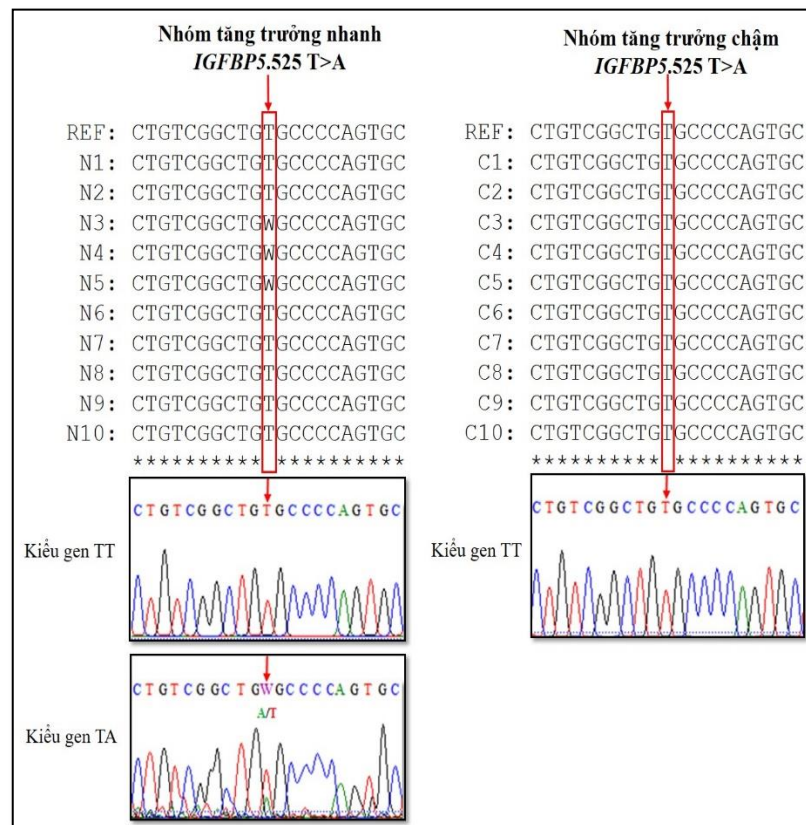
*: alen thiểu số; (^a): Tỷ lệ xuất hiện SNP của mỗi nhóm, ⁽¹⁾, ⁽²⁾: đáp ứng tiêu chí sàng lọc SNP số 1, 2 tương ứng.

Theo tiêu chí sàng lọc SNP, SNP 704C>G ở exon 1 nằm trong CDS làm codon CTC biến đổi thành GTC, gây thay đổi trình tự axit amin thứ 8 trong chuỗi polypeptit Leucine thành Valine, nên SNP này được sàng lọc để tiếp tục xác định trên bộ mẫu kiểm nghiệm (Hình 3.9), (Bảng 3.6). Các SNP còn lại nằm trong vùng không mã hóa được đánh giá tỷ lệ xuất hiện SNP cần đạt ít nhất 0,3 ở một nhóm trở lên, kết quả có 2 SNP

1123 G>T ở intron 1 và 17119 A>G ở exon 4 đạt tiêu chí này (Bảng 3.6). Tuy nhiên, 2 SNP này lại không thể hiện được sự khác biệt về thành phần kiểu gen hay thành phần alen giữa hai nhóm, với các giá trị p lớn hơn 0,05 (Bảng 3.6), nên chúng không được sàng lọc. Do đó, chỉ có SNP 704 C>G (p.Leu8Val) được sàng lọc để được tiếp tục xác định trên bộ mẫu lớn hơn.

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

Trình tự các đoạn gen *IGFBP-5* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu, từ đó phát hiện được 4 SNP thuộc các exon, trong đó có 2 SNP thuộc vùng mã hóa ở exon 1 và exon 4, 2 SNP thuộc vùng 3'-UTR ở exon 4 (Bảng 3.7). Định tính hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được minh họa ở Hình 3.10.



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.

Theo tiêu chí sàng lọc SNP, SNP 525T>A ở exon 1, nằm trong vùng mã hóa và gây ra sự thay đổi codon GTG thành GAG, làm thay đổi trình tự axit amin thứ 16 trong chuỗi peptit từ Valine thành axit Glutamic nên SNP này được sàng lọc để xác định trên bộ mẫu lớn hơn (Hình 3.10), (Bảng 3.7). Cũng nằm trong vùng mã hóa, nhưng SNP 8599 G>A không làm thay đổi trình tự axit amin, nên SNP đó cùng hai SNP còn lại nằm trong vùng không mã hóa được kiểm tra tỉ lệ xuất hiện SNP ở hai nhóm tăng trưởng nhanh/chậm. Kết quả chỉ có SNP 11992 T>C có tỉ lệ xuất hiện SNP đạt ít nhất 0,3 ở một nhóm trở lên (Bảng 3.7) và được tiếp tục kiểm tra sự khác biệt về thành phần kiểu gen/thành phần alen giữa hai nhóm, tuy nhiên với các giá trị p lớn hơn 0,05, SNP này không đạt được tiêu chí sàng lọc. Do đó, chỉ có SNP 525 T>A (p.Val16Glu) được sàng lọc để được tiếp tục được xác định trên bộ mẫu kiểm nghiệm.

Bảng 3.7. Danh sách các SNP được phát hiện trên gen IGFBP-5

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	E1_CDS_525 ⁽¹⁾	T	A*	7TT:3TA	10TT	17T:3A	20T
2	E4_CDS_8599	G	A*	8GG:2AG (0,20)	9GG:1GA (0,10)	18G:2A	19G:1A
3	E4_3'UTR_11713	C	A*	8CC:1CA:1AA (0,22)	10CC (0,00)	17C:3A	20C
4	E4_3'UTR_11992 ⁽²⁾	T	C*	6TT:1TC:3CC (0,57)	7TT:2TC:1CC (0,33)	13T:7C	16T:4C
				p = 0,55		p = 0,48	

*: alen thiếu số; (^a): Tỉ lệ xuất hiện SNP của mỗi nhóm, ⁽¹⁾, ⁽²⁾: đáp ứng tiêu chí sàng lọc SNP số 1, 2 tương ứng.

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

Trình tự các đoạn gen *IGFBP-6* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu để xác định SNP. Định tín hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được minh họa ở Hình 3.11. Có 17 SNP được phát hiện ở gen *IGFBP-6*, trong đó có 14 SNP thuộc các intron và 3 SNP nằm ở vùng 3'UTR của exon 4 (Bảng 3.8).

Bảng 3.8. Danh sách các SNP được phát hiện trên gen *IGFBP-6*

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	
2	I1_770	T	C*	8TT:1TC:1CC (0,22)	10TT (0,00)	17T:3C	20T
3	I2_1517	C	T*	8CC:2TT (0,25)	8CC:2TT (0,25)	16C:4T	16C:4T
4	I2_1527	C	T*	10CC (0,00)	8CC:2TT (0,25)	20C	16C:4T
5	I3_1898	T	C*	9TT:1TC (0,10)	8TT:2TC (0,20)	19T:1C	18T:2C
6	I3_2121	G	T*	10GG (0,00)	8GG:2TT (0,25)	20G	16G:4T
7	I3_2141	G	T*	9GG:1GT (0,10)	8GG:2GT (0,20)	19G:1T	18G:2T
8	I3_2198	T	C*	10TT (0,00)	8TT:2CC (0,25)	20T	16T:4C
9	I3_2225	G	T*	9GG:1TT (0,11)	8GG:2TT (0,25)	18G:2T	16G:4T
10	I3_2244	C	T*	9CC:1TT (0,11)	10CC (0,00)	18C:2T	20C
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	
14	I3_2671	A	T*	9AA:1TT (0,11)	10AA (0,00)	18A:2T	20A

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
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	

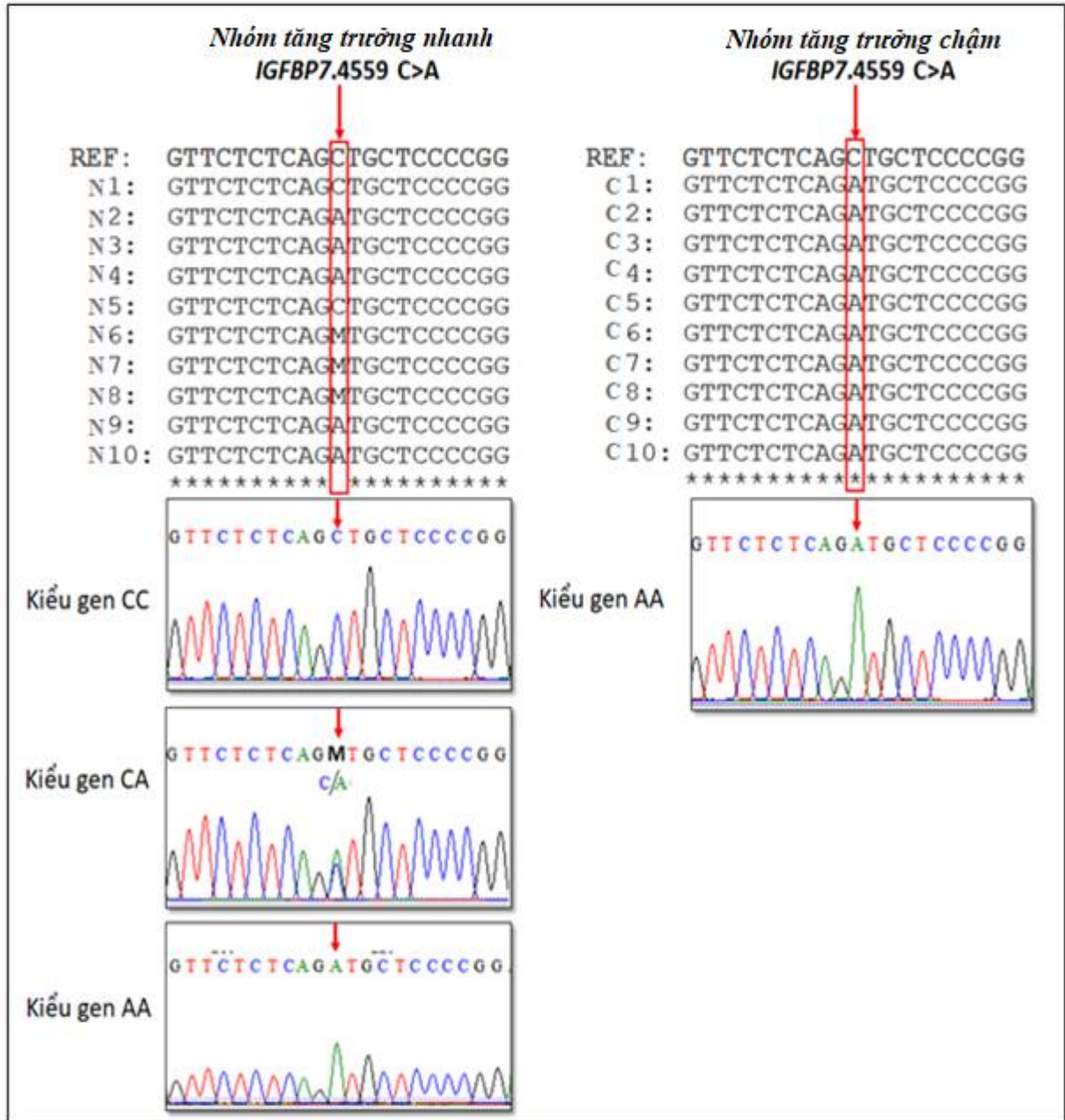
*: alen thiểu số; (^a): Tỷ lệ xuất hiện SNP của mỗi nhóm, (²), (⁴): đáp ứng tiêu chí sàng lọc SNP số 2, 4 tương ứng.

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

Trình tự các đoạn gen *IGFBP-7* từ 10 cá thể tăng trưởng nhanh và 10 cá thể tăng trưởng chậm được so sánh với trình tự tham chiếu để xác định SNP. Định tín hiệu giải trình tự của từng vị trí SNP được kiểm tra. Ví dụ về một vị trí SNP được minh họa ở Hình 3.12. Có 27 SNP được phát hiện trên gen *IGFBP-7*, trong đó có 25 SNP nằm trên vùng không mã hóa và 2 SNP gây nên sự thay thế axit amin là SNP 344T>C trên exon 1 và 4559C>A trên exon 3, lần lượt dẫn đến thay đổi trình tự axit amin Leu78Pro và Leu189Met trên protein (Bảng 3.9).

Theo tiêu chí sàng lọc, 2 SNP gây nên sự thay đổi axit amin là 344 T>C (p.Leu78Pro) và 4559 C>A (p.Leu189Met) sẽ được sàng lọc để genotyping ở bộ mẫu lớn hơn. Trong khi đó, các SNP còn lại sẽ được đánh giá tỷ lệ xuất hiện SNP ở hai nhóm tăng trưởng nhanh/chậm phải đạt ít nhất 0,3 ở một nhóm trở lên. Kết quả được trình bày ở Bảng 3.9 cho thấy có 17 SNP đạt tiêu chí này là -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 (Bảng 3.9). Trong số 17 SNP này, chỉ có SNP 2060 A>G đạt tiêu chí về sự khác biệt có ý nghĩa về thành phần kiểu gen, cũng như thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm (với cả 2 giá trị p nhỏ hơn 0,05) (Bảng

3.9). Do vậy, SNP 2060 A>G trên intron 2 và 2 SNP 344 T>C (p.Leu78Pro) và 4559 C>A (p.Leu189Met) thuộc vùng mã hóa gây thay đổi trình tự axit amin sẽ được sàng lọc để tiếp tục genotyping ở bộ mẫu lớn hơn.



Hình 3.12. Kết quả phát hiện IGFBP-7.4559 C>A trên gen IGFBP-7 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.

Bảng 3.9. Danh sách các SNP được phát hiện trên gen *IGFBP-7*

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN (^a)	TTC (^a)	TTN	TTC
1	Promoter -376 ⁽²⁾	T	G*	5TT:3TG:2GG (0,63)	4TT:1TG:5GG (1,20)	13T:7G	9T:11G
				p = 0,39		p = 0,34	
2	E1_CDS 344 ⁽¹⁾	T*	C	1TT:2TC:7CC	3TT:3TC:4CC	4T:16C	9T:11C
3	I1_561 ⁽²⁾	G	A*	6GG:1GA:3AA (0,57)	2GG:3GA:5AA (1,60)	13G:7A	7G:13A
				p = 0,19		p = 0,11	
4	I1_598 ⁽²⁾	C	A*	10CC (0,00)	7CC:2CA:1AA (0,33)	20C	16C:4A
				p = 0,11		p = 0,10	
5	I1_1052 ⁽²⁾	T*	A	4TT:1TA:5AA (0,83)	3TT:1TA:6AA (0,57)	9T:11A	7T:13A
				p = 1,0		p = 0,75	
6	I1_1117 ⁽²⁾	C	G*	7CC:2CG:1GG (0,33)	5CC:2CG:3GG (0,71)	16C:4G	12C:8G
				p = 0,51		p = 0,30	
7	I1_1160 ⁽²⁾	G	C*	7GG:2GC:1CC (0,33)	10GG (0,00)	16G:4C	20G
				p = 0,11		p = 0,10	
8	I1_1216 ⁽²⁾	C	A*	5CC:2CA:3AA (0,71)	8CC:1CA:1AA (0,22)	12C:8A	17C:3A
				p = 0,36		p = 0,16	
9	I1_1597	C	G*	8CC:2CG (0,20)	10CC (0,00)	18C:2G	20C
10	I1_1618	T	C*	9TT:1CC (0,11)	9TT:1CC (0,11)	18T:2C	18T:2C
11	I2_2030	A	G*	9AA:1AG (0,10)	10AA (0,00)	19A:1G	20A
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	
13	I2_ 21382 ⁽²⁾	T*	C	1TC:9CC (0,10)	3TC:7CC (0,30)	1T:19C	3T:17C
				p = 0,58		p = 0,60	
14	I2_2240	T	C*	9TT:1CC (0,11)	10TT (0,00)	18T:2C	20T

STT	Vị trí SNP trên gen	Ref	Alt	Thành phần kiểu gen		Thành phần alen	
				TTN ^(a)	TTC ^(a)	TTN	TTC
15	I2_2634	C	T*	9CC:1TT (0,11)	8CC:2CT (0,20)	18C:2T	18C:2T
16	I2_3981 ⁽²⁾	T*	G	4TT:6GG (0,67)	3TT:7GG (0,43)	8T:12G	6T:14G
				p = 1,0		p = 0,74	
17	I2_4058 ⁽²⁾	A*	G	4AA:1AG:5GG (0,83)	4AA:2AG:4GG (1,00)	9A:11G	10A:10G
				p = 0,86		p = 0,99	
18	I2_4334 ⁽²⁾	G	A*	7GG:2GA:1AA (0,33)	8GG:1GA:1AA (0,22)	16G:4A	17G:3A
				p = 0,79		p = 1,0	
19	I2_4465 ⁽²⁾	C	T*	4CC:4CT:2TT (0,75)	1CC:7CT:2TT (1,13)	12C:8T	9C:11T
				p = 0,30		p = 0,53	
20	E3_CDS4 559⁽¹⁾	C*	A	2CC:3CA:5AA	10AA	7C:13A	20A
21	I3_4685 ⁽²⁾	T	C*	4TT:1TC:5CC (1,20)	7TT:3CC (0,43)	9T:11C	14T:6C
				p = 0,26		p = 0,20	
22	I3_4761	A	G*	9AA:1AG (0,01)	10AA (0,00)	19A:1G	20A
23	I4_7405 ⁽²⁾	C	T*	8CC:1CT:1TT (0,22)	6CC:2CT:2TT (0,50)	17C:3T	14C:6T
				p = 0,51		p = 0,45	
24	I4_7486	G	A*	10GG (0,00)	9GG:1GA (0,10)	20G	19G:1A
25	I4_7676 ⁽²⁾	G	C*	7GG:3CC (0,43)	5GG:5CC (1,00)	14G:6C	10G:10C
				p = 0,65		p = 0,33	
26	I4_7692 ⁽²⁾	T	C*	6TT:4CC (0,67)	5TT:5CC (1,00)	12T:8C	10T:10C
				p = 1,00		p = 0,75	
27	I4_8154	A*	T	1AT:9TT (0,10)	10TT (0,00)	1A:19T	20T

*: alen thiểu số; (a): Tỷ lệ xuất hiện SNP của mỗi nhóm, (1), (2), (3), (4): đáp ứng tiêu chí sàng lọc SNP số 1, 2, 3, 4 tương ứng.

Nhìn chung, luận án đã phát hiện được 10, 12, 9, 4, 10, 10, 4, 17, 27 SNP lần lượt trên các gen *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7 thuộc 20 cá thể của bộ mẫu khởi tạo, trong đó có 10 SNP được sàng lọc theo các tiêu chí đã đề ra để tiếp tục được xác định trên bộ mẫu lớn hơn (Bảng 3.10).

Bảng 3.10. Danh sách các SNP được sàng lọc trên bộ mẫu khởi tạo

STT	Tên gen	Vị trí SNP trên gen	Ref	Alt
1	<i>IGF1</i>	Intron 2_13680	A	T
2	<i>IGF1R</i>	Intron 1_13357	T	C
3		Intron 1_15392	T	A
4		Intron 15_83894	A	G
5	<i>IGFBP-3</i>	Exon 1_CDS_704	C	G
6	<i>IGFBP-5</i>	Exon 1_CDS_525	T	A
7	<i>IGFBP-6</i>	Intron 3_2278	C	A
8	<i>IGFBP-7</i>	Exon 1_CDS_344	T	C
9		Intron 2_2060	A	G
10		Exon 3_CDS_4559	C	A

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

3.4.1.1. Phân lập các đoạn gen có chứa SNP được sàng lọc

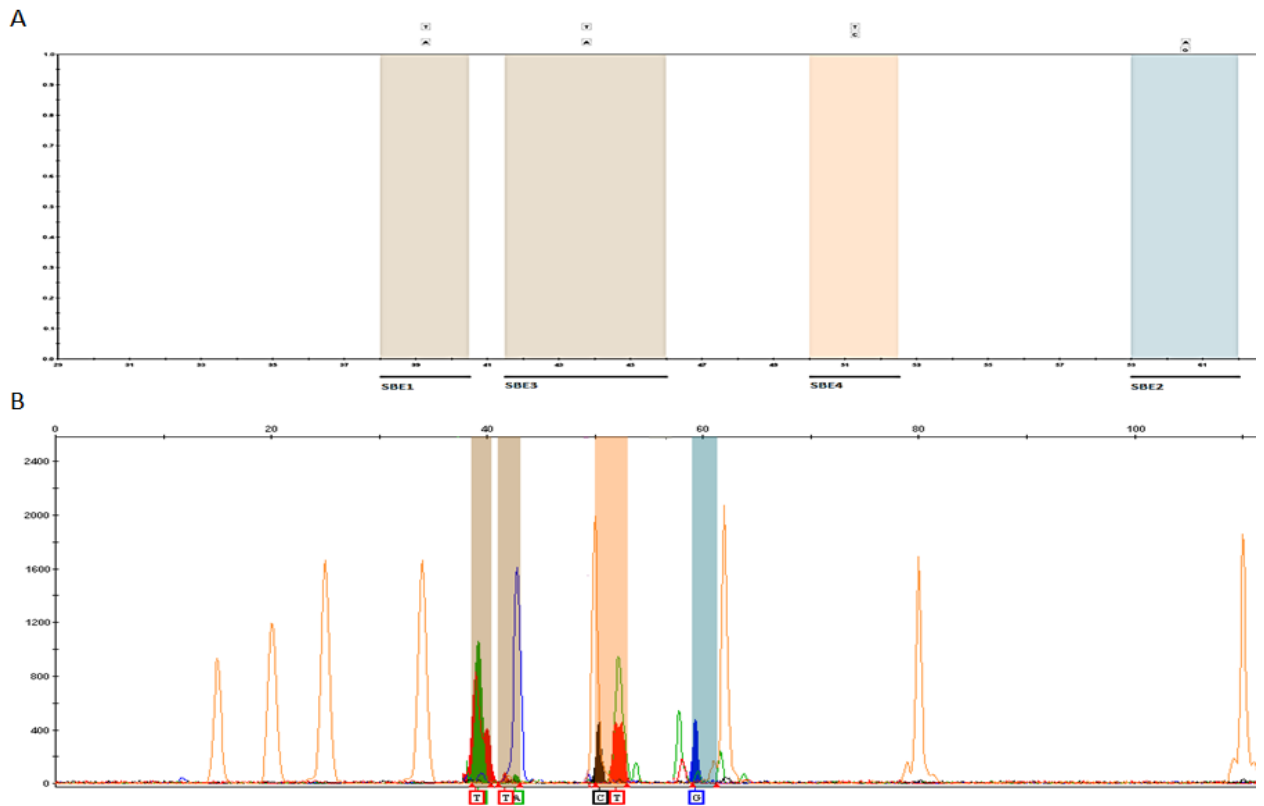
Một trăm bốn mươi mẫu DNA tổng số được tách chiết từ 70 cá thể cá tra tăng trưởng nhanh và 70 mẫu cá tra tăng trưởng chậm (thông tin mẫu trình bày tại Phụ lục 1) được điện di kiểm tra chất lượng trên gel agarose 1% và đo nồng độ và độ tinh sạch bằng máy đo quang phổ ở bước sóng A260 và A280. Kết quả điện di một số mẫu (trình bày ở phần Phụ lục 6) cho thấy các mẫu đều có chất lượng tốt với một băng DNA tổng số duy nhất, rõ nét, không bị đứt gãy và kết quả đo nồng độ, độ tinh sạch (trình bày ở Phụ lục 7) cho thấy tất cả các mẫu DNA tổng số đều đủ tiêu chuẩn để sử dụng cho các thí nghiệm tiếp theo.

Kết quả điện di sản phẩm PCR khuếch đại các đoạn gen chứa SNP được sàng lọc trên 140 cá thể thuộc bộ mẫu kiểm nghiệm được trình bày trong Phụ lục 8, cho thấy các

băng thu được rõ nét, đặc hiệu với kích thước đúng như dự kiến, đủ tiêu chuẩn để được tinh sạch và sử dụng cho phản ứng genotyping kéo dài một nucleotide.

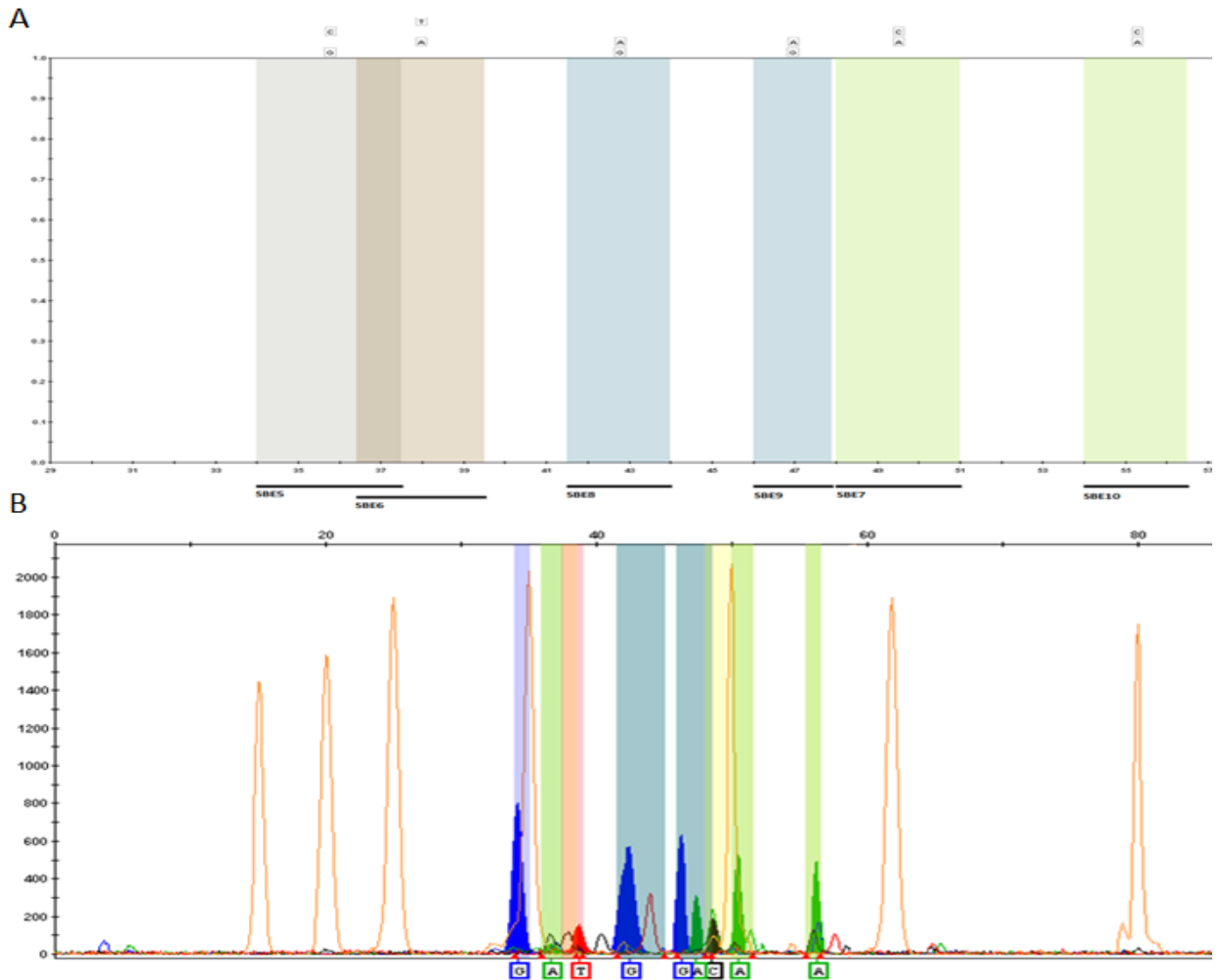
3.4.1.2. Xác định kiểu gen của từng cá thể ở vị trí SNP

Sản phẩm PCR khuếch đại các đoạn gen chứa 10 SNP được sàng lọc sau tinh sạch của 70 cá thể tăng trưởng nhanh và 70 cá thể tăng trưởng chậm được sử dụng làm khuôn cho phản ứng kéo dài một nucleotide multiplex nhằm genotyping các SNP đó theo hai nhóm môi (Bảng 2.11). Kết quả thô thu được từ máy giải trình tự của hai nhóm phản ứng SBE trên một cá thể được trình bày ở Hình 3.13 và Hình 3.14, cho thấy các sản phẩm có chiều dài khác nhau được phân tách rõ trên điện di mao quản, đỉnh màu huỳnh quang ứng với mỗi vùng vị trí liên quan đến mỗi SNP rõ ràng cho phép nhận biết và genotyping chính xác được SNP đó.



Hình 3.13. Kết quả điện di mao quản sản phẩm SNaPshot trên một mẫu đại diện của nhóm môi SBE thứ nhất sau khi được phân tích trên phần mềm GeneMapper.

A. Binset mô tả chiều dài của các sản phẩm ứng với mỗi SNP trong nhóm được hiển thị trong điện di mao quản. B. Các đỉnh huỳnh quang của các SNP trong nhóm được hiển thị sau khi chạy phần mềm phân tích kết quả GeneMapper.



Hình 3.14. Kết quả điện di mao quản SNaPshot trên một mẫu đại diện của nhóm môi SBE thứ hai sau khi được phân tích trên phần mềm GeneMapper.

A. Binset mô tả chiều dài của các sản phẩm thu được ứng với mỗi SNP trong nhóm được hiển thị trong điện di mao quản. B. Các đỉnh huỳnh quang của các SNP trong nhóm được hiển thị sau khi chạy phần mềm phân tích kết quả GeneMapper.

Kết quả genotyping 10 SNP trên 140 cá thể, kết hợp cùng với dữ liệu SNP thu được từ bộ mẫu khởi tạo gồm 20 cá thể được trình bày trong Phụ lục 9, cho thấy số lượng cá thể được xác định kiểu gen đạt yêu cầu cho các phép phân tích dữ liệu.

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

Kết quả SNP genotyping trên bộ mẫu kiểm nghiệm kết hợp với kết quả phát hiện và sàng lọc SNP trên bộ mẫu khởi tạo trước đó được sử dụng để phân tích mối liên

quan giữa 10 SNP được sàng lọc với tính trạng tăng trưởng của cá tra nuôi trên 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm thông qua: thành phần kiểu gen, thành phần alen; các thông số về quần thể gồm PIC, MAF; trạng thái mất cân bằng liên kết và tỉ lệ các kiểu gen đơn bội haplotype tạo nên từ các SNP được sàng lọc trên cùng một gen; và tỉ lệ các kiểu gen đơn bội haplotype, tổ hợp kiểu gen tạo nên từ các SNP tiềm năng trên các gen khác nhau.

3.4.2.1. SNP được sàng lọc trên gen *IGF1*

Thành phần kiểu gen và thành phần alen của 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm cùng các chỉ số di truyền quần thể từ SNP 13680 A>T trên gen *IGF1* được trình bày trong Bảng 3.11.

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, **: $p < 0,01$.

Khi so sánh với bộ mẫu khởi tạo (Bảng 3.1), thành phần kiểu gen phát sinh từ SNP *IGF1*:13680 A>T ở bộ mẫu lớn hơn của nhóm tăng trưởng nhanh có thêm kiểu gen AT. Tuy nhiên, kiểu gen AA vẫn chiếm tỉ lệ lớn hơn kiểu gen TT ở nhóm tăng trưởng nhanh và kiểu gen TT ngang bằng với kiểu gen AA ở nhóm tăng trưởng chậm (Bảng 3.11). Đáng chú ý, mặc dù không còn thể hiện được sự khác biệt về thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm (với p bằng 0,06), SNP này vẫn biểu hiện được sự khác nhau rõ rệt về thành phần kiểu gen giữa hai nhóm với p nhỏ hơn 0,01 (p bằng 0,009) khi được phân tích trên bộ mẫu lớn hơn (Bảng 3.11). Như vậy, SNP 13680 A>T ở intron

2 của gen *IGF1* liên quan đến tính trạng tăng trưởng của cá tra nuôi. Ngoài ra SNP này có giá trị PIC là 0,494, nằm trong khoảng 0,25 đến 0,5, cho thấy SNP này có mức độ đa dạng di truyền trung bình. Giá trị MAF của SNP này là 0,446, cho thấy SNP 13680 A>T xuất hiện phổ biến trong quần thể.

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

a) Thành phần kiểu gen, thành phần alen, các chỉ số PIC, MAF

Thành phần kiểu gen và thành phần alen của 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm cùng các chỉ số di truyền quần thể từ 3 SNP 13357 T>C, 15392 T>A và 83894 A>G trên gen *IGF1R* được trình bày trong 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
	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	p	Nhóm tăng trưởng nhanh	Nhóm tăng trưởng chậm	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

^(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, *: $p < 0,05$.

Dữ liệu về thành phần kiểu gen của 3 SNP trên thu được từ 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm (Bảng 3.12) có nhiều kiểu gen được xác định hơn so với bộ mẫu khởi tạo (Bảng 3.3). Trong số đó, chỉ còn SNP 13357 T>C tiếp tục 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), với sự xuất hiện của kiểu gen CC ở nhóm tăng trưởng chậm cao hơn so với nhóm tăng trưởng nhanh (Bảng 3.12). Như vậy, SNP 13357 T>C trên intron 1 của gen này được coi là chỉ thị MAS tiềm năng hỗ trợ chọn lọc tăng trưởng cho cá tra nuôi.

Các giá trị PIC của cả 3 SNP đều nằm trong khoảng 0,25 đến 0,5 cho thấy chúng đều có mức độ đa dạng di truyền ở mức độ trung bình. Các giá trị MAF của cả 3 SNP đều xấp xỉ 0,5 cho thấy chúng đều là cả 3 SNP đều là các biến dị xuất hiện thường xuyên trong quần thể.

b) Sự mất cân bằng liên kết thông qua giá trị D' và R^2

Sự mất cân bằng liên kết được sử dụng để đánh giá khả năng di truyền cùng nhau của 3 SNP trên thông qua giá trị D' và R^2 . Dữ liệu trên tổng số 160 cá thể của 3 SNP trên (Phụ lục 9) được sử dụng để khảo sát sự mất cân bằng liên kết này. Với giá trị D' cao 0,935 (D' lớn hơn 0,8), SNP 15392 T>A và SNP 83894 A>G có khả năng cao di truyền cùng nhau; với giá trị D' trung bình 0,718 và 0,490 (D' nằm trong khoảng từ 0,33 đến 0,80), SNP 15392 T>A và SNP 13357 T>C, SNP 83894 A>G và SNP 13357 T>C có khả năng di truyền cùng nhau (Bảng 3.13). Các giá trị R^2 giữa các cặp SNP đều lớn hơn 0 (Bảng 3.13), cũng cho thấy 3 SNP trên gen *IGF1R* đạt trạng thái mất cân bằng liên kết, có khả năng cao được di truyền cùng nhau.

Bảng 3.13. Phân tích sự mất cân bằng liên kết (linkage disequilibrium (LD)) của 3 SNP trên gen *IGF1R* thông qua giá trị D' và R^2 .

D' và R^2	13357 T>C	15392 T>A	83894 A>G
13357 T>C	- và -	0,718 và 0,223	0,490 và 0,124
15392 T>A	0,718 và 0,223	- và -	0,935 và 0,743
83894 A>G	0,490 và 0,124	0,935 và 0,743	- và -

c) Tần số kiểu gen đơn bội (haplotype)

Sự khác biệt về thành phần và tỉ lệ các haplotype tạo nên từ 3 SNP 13357 T>C, 15392 T>A và 83894 A>G ở nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm cũng

được phân tích từ kết quả genotyping của 3 SNP trên tổng số 160 cá thể (Phụ lục 9), 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). Haplotype H5 với 42,3% và H3 với 36,5% lần lượt chiếm tỉ lệ lớn nhất trong tổng số các haplotype ở nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm (Bảng 3.14). Các kết quả trên này cho thấy các haplotype được hình thành từ 3 SNP trên gen *IGF1R* có mối liên quan chặt chẽ với tính trạng tăng trưởng của cá tra nuôi.

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***

^(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 và thành phần alen của 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm cùng các chỉ số di truyền quần thể từ SNP 704 C>G (p.Leu8Val) trên gen *IGFBP-3* được trình bày trong Bảng 3.15.

Theo đó, chỉ có 2 kiểu gen phát sinh từ SNP này ở cả hai nhóm, tuy nhiên, ở nhóm tăng trưởng nhanh, kiểu gen CG chiếm đa số, trong khi ở nhóm tăng trưởng chậm, kiểu gen GG chiếm ưu thế (Bảng 3.15). Sự khác biệt về thành phần kiểu gen giữa hai nhóm tăng trưởng nhanh/chậm có ý nghĩa, với giá trị p là 0,03 ($p < 0,05$) (Bảng 3.15). Ngoài

ra, thành phần kiểu gen giữa hai nhóm cũng khác biệt có ý nghĩa thống kê, với giá trị p là 0,013 ($p < 0,05$), cụ thể là tỉ lệ alen thiểu số C trên alen đa số G ở nhóm tăng trưởng nhanh là 45/115 (0,39) cao hơn hẳn tỉ lệ này ở nhóm tăng trưởng chậm là 26/132 (0,19) (Bảng 3.15). Như vậy, SNP 704 C>G (p.Leu8Val) trên gen *IGFBP-3* được coi là chỉ thị phân tử tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

Chỉ số PIC, MAF của SNP 704 C>G (p.Leu8Val) lần lượt là 0,347 và 0,223 (Bảng 3.15), tuy thấp hơn so với các SNP được sàng lọc ở các gen *IGF1* và *IGF1R* đã nêu trước đó, nhưng vẫn thể hiện mức độ đa dạng di truyền và mức độ xuất hiện phổ biến của SNP này.

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, * $p < 0,05$, ** $p < 0,01$.

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

SNP 525 T>A (p.Val16Glu) được kiểm nghiệm sự liên quan đến tính trạng tăng trưởng của cá tra nuôi thông qua sự khác biệt về thành phần kiểu gen và/ hoặc thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, cũng như đánh giá mức độ đa dạng di truyền thông qua chỉ số di truyền quần thể PIC, MAF (Bảng 3.16).

Kết quả cho thấy sự khác biệt có ý nghĩa về cả thành phần kiểu gen lẫn thành phần alen phát sinh từ SNP này giữa hai nhóm tăng trưởng nhanh/chậm (với cả 2 giá trị p đều nhỏ hơn 0,001) (Bảng 3.16). Cụ thể, nhóm tăng trưởng nhanh có 3 kiểu gen TT, TA và AA, với kiểu gen TA chiếm ưu thế, trong khi đó, nhóm tăng trưởng chậm chỉ có 2 kiểu

gen, với kiểu gen TT chiếm đa số. Về thành phần alen, trong khi số lượng alen T và alen A gần như ngang nhau ở nhóm tăng trưởng nhanh, số lượng alen T lại nhiều gấp 3 lần số lượng alen A ở nhóm tăng trưởng chậm (Bảng 3.16). Do đó, SNP 525 T>A (p.Val16Glu) được coi là chỉ thị phân tử tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi. Các chỉ số đa dạng di truyền PIC, MAF cho thấy SNP này có mức độ đa dạng di truyền trung bình và có mức độ xuất hiện phổ biến trong quần thể (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 intron 3 của gen *IGFBP-6* được kiểm nghiệm sự liên quan đến tính trạng tăng trưởng của cá tra nuôi thông qua sự khác biệt về thành phần kiểu gen và/ hoặc thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, cũng như đánh giá mức độ đa dạng di truyền thông qua chỉ số di truyền quần thể PIC, MAF. Tuy nhiên kết quả được trình bày ở Bảng 3.17 cho thấy không có sự khác biệt có ý nghĩa giữa thành phần kiểu gen và cả thành phần alen phát sinh từ SNP này giữa hai nhóm tăng trưởng nhanh/chậm, với các giá trị p đều lớn hơn 0,05 (Bảng 3.17). Các chỉ số đa dạng di truyền quần thể cho thấy SNP 2278 C>A của gen *IGFBP-6* có mức độ đa dạng di truyền trung bình và xuất hiện phổ biến trong quần thể.

Bảng 3.17. Phân tích mối liên quan giữa SNP 2278 C>A trên gen IGFBP-6 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)	CC (08) CA (57) AA (15)	CC (07) CA (55) AA (17) NN (01)	0,895
Thành phần alen ^(c)	C (73) A (87)	C (69) A (89)	0,726
PIC	0,494		
MAF	0,447		

^(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, * $p < 0,05$, ** : $p < 0,01$.

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

a) Thành phần kiểu gen, thành phần alen, các chỉ số PIC, MAF

Thành phần kiểu gen và thành phần alen của 80 cá thể tăng trưởng nhanh và 80 cá thể tăng trưởng chậm cùng các chỉ số di truyền quần thể từ 3 SNP 344 T>C (p.Leu78Pro), 2060 A>G và 4559 A>C (p.Leu189Met) trên gen *IGFBP-7* được trình bày trong Bảng 3.18.

Khi được phân tích trên 160 cá thể, SNP 344 T>C, vốn được sàng lọc do làm thay đổi axit amin trên protein, không thể hiện được sự khác nhau về thành phần kiểu gen, hay thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, với các giá trị p đều lớn hơn 0,05 (Bảng 3.18). Trong khi đó, SNP 2060 A>G trên intron 2 vẫn duy trì được sự khác biệt có ý nghĩa về cả thành phần kiểu gen và thành phần alen giữa hai nhóm tăng trưởng nhanh/chậm, với cả hai giá trị p đều nhỏ hơn 0,001 (Bảng 3.18). Cụ thể, ở nhóm tăng trưởng nhanh, kiểu gen GG và alen G chiếm tỉ lệ cao hơn hẳn kiểu gen AA và alen A, trong khi ở nhóm tăng trưởng chậm, kiểu gen AA và alen A lại chiếm tỉ lệ cao hơn kiểu gen GG và alen G (Bảng 3.18).

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, *: $p < 0,05$ **: $p < 0,01$, ***: $p < 0,001$.

Khác với SNP 344 T>C, SNP 4559 C>A gây sự thay thế axit amin trên exon 3 vẫn tiếp tục thể hiện được sự khác biệt về thành phần kiểu gen giữa hai nhóm, với kiểu gen AC ở nhóm tăng trưởng chậm cao hơn ở nhóm tăng trưởng nhanh (p nhỏ hơn 0,05) (Bảng 3.18). Như vậy, 2 SNP 2060 A>G trên intron 2 và SNP 4559 C>A (p.Leu189Met) trên exon 3 của gen *IGFBP-7* được coi là 2 chỉ thị SNP tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

Về chỉ số đa dạng di truyền quần thể của các SNP (Bảng 3.18), trong khi SNP 344 T>C (p.Leu78Pro) và 2060 A>G thể hiện sự mức độ đa dạng di truyền trung bình với các giá trị PIC thuộc khoảng từ 0,25 đến 0,5 và mức độ xuất hiện phổ biến trong quần thể với các giá trị MAF đều lớn hơn 0,2, SNP 4559 C>A lại có 2 giá trị PIC (0,216) và MAF (0,123) nhỏ hơn các khoảng giá trị này (Bảng 3.18), cho thấy mức độ đa dạng di truyền thấp cũng như sự kém phổ biến trong quần thể của nó.

b) Sự mất cân bằng liên kết thông qua giá trị D' và R^2

Sự mất cân bằng liên kết được sử dụng để đánh giá khả năng di truyền cùng nhau của 3 SNP trên thông qua giá trị D' và R^2 . Dữ liệu trên tổng số 160 cá thể của 3 SNP trên (Phụ lục 9) được sử dụng để khảo sát sự mất cân bằng liên kết này. Theo đó chỉ có SNP 344 T>C và 4559 C>A có giá trị D' rất cao bằng 0,999 (Bảng 3.19). Các cặp SNP 344 T>C và 2060 A>G, 2060 A>G và 4559 C>A đều có cả giá trị D' và R^2 nhỏ hơn 0,33 (Bảng 3.19). Tuy thể hiện trạng thái mất cân bằng liên kết ở mức độ yếu, 3 SNP vẫn có khả năng di truyền cùng nhau.

Bảng 3.19. Phân tích sự mất cân bằng liên kết (linkage disequilibrium (LD)) của 3 SNP trên gen *IGFBP-7* thông qua giá trị D' và R^2 .

D' và R^2	344 T>C	2060 A>G	4559 C>A
344 T>C	- và -	0,210 và 0,024	0,999 và 0,073
2060 A>G	0,210 và 0,024	- và -	0,231 và 0,007
4559 C>A	0,999 và 0,073	0,231 và 0,007	- và -

c) Tần số kiểu gen đơn bội (haplotype)

Mối tương quan giữa haplotype được tạo nên từ 3 SNP của gen *IGFBP-7* với tính trạng tăng trưởng vẫn được phân tích trong luận án (Bảng 3.20) cho thấy, trong số 6 haplotype, haplotype *IGFBP-7_H3* (CGA) chiếm tỉ lệ cao nhất ở nhóm tăng trưởng nhanh, còn ở nhóm tăng trưởng chậm, tỉ lệ cao nhất thuộc về haplotype H1 (CAA). Ngoài ra, 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 (p bằng 2,97e-7 và 3,97e-5). Trong khi đó, haplotype *IGFBP-7_H3* (CGA) và H4 (CGC) lại tương quan chặt chẽ với nhóm tăng trưởng nhanh với giá trị p nhỏ hơn 0,001 (p bằng 1,31e-7) và p nhỏ hơn 0,01 (p bằng 0,0054) (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

^(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.

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

a) Tỷ lệ haplotype tạo nên từ 6 SNP tiềm năng

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.Val116Glu), *IGFBP-7*. 2060 A>G và *IGFBP-7*. 4559 C>A (p.Leu189Met) đã được phân tích thông qua sự khác biệt về tỷ lệ haplotype cũng như tổ hợp kiểu gen giữa hai nhóm tăng trưởng nhanh và tăng trưởng chậm (p < 0,05).

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
IGF1 13680 A>T	IGF1R 13357 T>C	IGF BP3 704 C>G	IGF BP5 525 T>A	IGF BP7 2060 A>G	IGF BP7 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, ^(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<0,05, ** p<0,01, *** p<0,001.

Phân tích tác động của các haplotype phát sinh từ 6 chỉ thị SNP trên lên tính trạng tăng trưởng của cá tra nuôi (Bảng 3.21) cho thấy có 4 haplotype chiếm tỉ lệ cao hơn ở

nhóm tăng trưởng nhanh là CO_H2, H5, H11, H14 (với các giá trị p đều < 0,05) và 7 haplotype chiếm tỉ lệ cao hơn ở nhóm tăng trưởng chậm là CO_H1, H3, H7, H8, H9, H10, H14 (với các giá trị p đều < 0,05). Trong đó, haplotype CO_H2 chiếm tỉ lệ cao nhất trong nhóm tăng trưởng nhanh (0,414) và cho thấy sự khác biệt rõ rệt nhất về tần số giữa nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm ($p=2,22e-16$) (Bảng 3.21). Như vậy, các chỉ thị haplotype CO_H2 (ATGTGA), CO_H5 (TTCAGA), CO_H11 (TCCAAA) và CO_H14 (TCGAAA) phát sinh từ 6 chỉ thị SNP trên có thể được coi như các chỉ thị phân tử hỗ trợ chọn giống cá tra theo hướng tăng trưởng nhanh.

b) Tỉ lệ tổ hợp kiểu gen tạo nên từ 6 SNP tiềm năng

Ứng với các vị trí SNP *IGF1*. 13680 A>T, *IGF1R*. 13357 T>C, *IGFBP-3*. 704C>G, *IGFBP-5*. 525T>A, *IGFBP-7*. 2060 A>G và *IGFBP-7*. 4559 C>A, 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).

Kết quả này phù hợp với kết quả về thành phần kiểu gen của mỗi chỉ thị SNP trên 160 cá thể, cụ thể với SNP *IGF1*.13680 A>T, tần số kiểu gen AT ngang nhau ở hai nhóm tăng trưởng nhanh/chậm (Bảng 3.11), với SNP *IGF1R*. 13357 T>C tần số kiểu gen CT cũng như TT ngang nhau ở hai nhóm tăng trưởng nhanh/chậm (Bảng 3.12), với SNP *IGFBP-3*. 704 C>G tần số kiểu gen GC chiếm tỉ lệ cao hơn ở nhóm tăng trưởng nhanh (Bảng 3.15), với SNP *IGFBP-5*. 525 T>A tần số kiểu gen AT chiếm tỉ lệ cao hơn ở nhóm tăng trưởng nhanh (Bảng 3.16), với SNP *IGFBP-7*. 2060 A>G tần số kiểu gen GA cũng như GG chiếm tỉ lệ cao hơn ở nhóm tăng trưởng nhanh và với SNP *IGFBP-7*. 4559 C>A tần số kiểu gen AA ngang nhau ở hai nhóm tăng trưởng nhanh/chậm (Bảng 3.18). Hai tổ hợp kiểu gen đại diện cho nhóm tăng trưởng nhanh trên được xác định diplotype là H2/H11 và H2/H5, được tạo nên từ những haplotype H2, H5, H11 cũng đều là những haplotype đại diện cho nhóm tăng trưởng nhanh (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, * $p < 0,05$, ** $p < 0,01$, *** $p < 0,001$.

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

Tất cả các gen đích thuộc hệ thống IGF của cá tra nuôi được phân tích trong luận án đều có mức độ tương đồng về trình tự cao nhất với các trình tự tương ứng của cá nheo Mỹ (*Ictalurus punctatus*) (Kết quả được nêu ở các phần thuộc mục 3.1). Đây là một kết quả hợp lý vì cá nheo Mỹ là một loài rất gần với cá tra nuôi trong hệ thống phân loại, cùng thuộc bộ cá da trơn hay cá nheo (*Siluriformes*). Genome của một trong những loài cá chủ lực trong nuôi trồng thủy sản của Mỹ này lần đầu tiên được công bố vào năm 2016, với ước tính khoảng 54.000 SNP và khoảng 26.661 gen mã hóa protein [92].

Trình tự protein của các gen thuộc hệ thống IGF được quan tâm trong luận án được so sánh với các trình tự tương ứng của các loài cá xương khác để tìm ra các vùng bảo thủ, cũng như các đoạn trình tự quan trọng liên quan đến chức năng của các protein, nhằm đánh giá ban đầu các tác động của các SNP gây thay đổi trình tự axit amin được tìm thấy khi phân tích đa hình các gen. Các so sánh đó được trình bày trong Phụ lục 10, cho thấy protein IGF1 của cá tra nuôi có sự tương đồng về mặt cấu trúc các domain B, C, A, D và E từ đầu N đến đầu C với protein các loài cá khác, với mức độ bảo thủ cao ở domain A, B [93] và C [32] chứa các trình tự thuộc motif điển hình bám với thụ thể tuýp 1, 2 của insulin/IGF và các IGFBP [32], [94], [95]. Domain D của các IGF1 của cá được cho là có độ bảo thủ ít nhất so với các domain khác, với độ tương đồng chỉ đạt 37,5 đến 75%, trong khi tỉ lệ này ở domain E là từ 62,16 đến 100% [93]. Điều này cũng được nhận thấy khi phân tích cấu trúc IGF1 của cá tra nuôi, có thể do domain E của IGF1 đóng vai trò quan trọng trong việc vận chuyển nội bào, hoàn thiện và gấp nếp protein [32] và các biến dị trong cấu trúc của domain E trong IGF1 có thể liên quan đến thời gian tồn tại của IGF1 trong hệ tuần hoàn [93].

Khác với IGF1 có vùng domain D với trình tự nhiều biến dị, vùng D ở IGF2 lại rất bảo thủ [32], điều này cũng được nhận thấy trên domain D của IGF2 cá tra nuôi thông qua so sánh với trình tự protein IGF2 của các loài cá xương khác (Phụ lục 11). Domain E-phần sẽ bị loại bỏ khỏi IGF2 để khiến protein này được hoàn thiện, có hoạt tính sinh học, thường có độ dài lớn hơn 98 axit amin [96]. Trong nghiên cứu này, độ dài của

domain E là 101 axit amin và chứa IGF2_C_super family tương đồng với trình tự của các loài cá xương khác (Phụ lục 11).

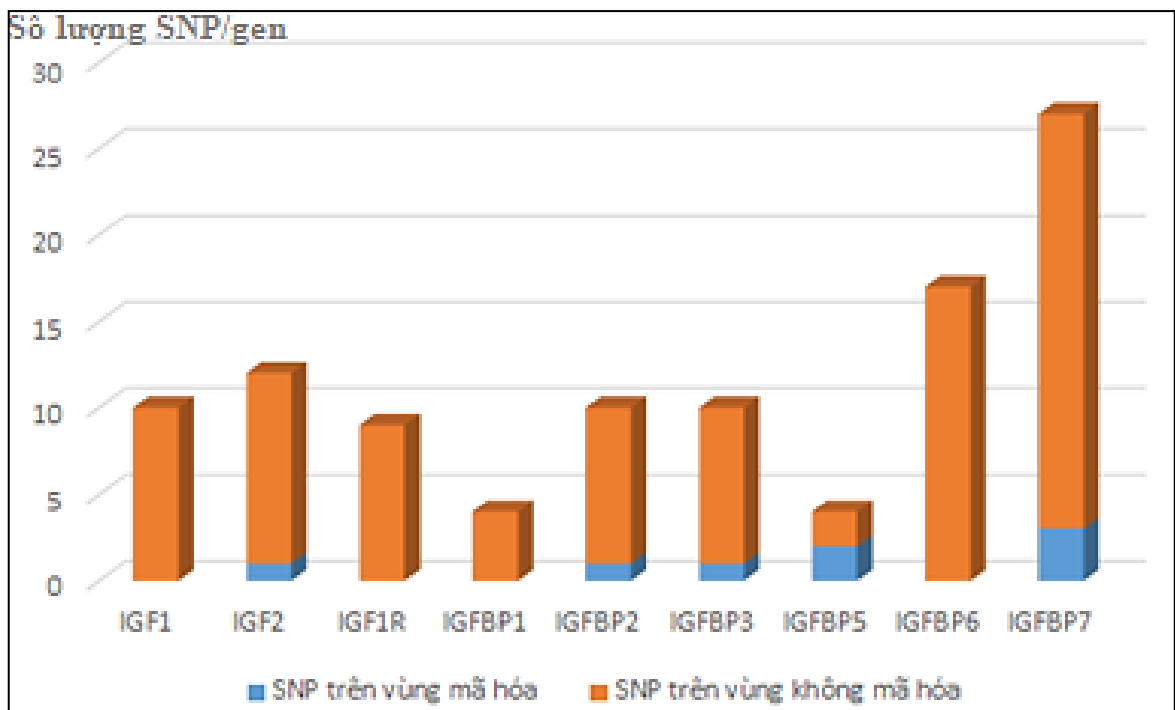
Các motif và các vị trí điển hình trong IGF1Ra của cá ngựa vằn (*D. rerio*) [57] đã được tìm thấy trong cấu trúc IGF1R của cá tra nuôi (Phụ lục 12), bao gồm: vị trí thủy phân gồm 4 axit amin cơ bản RRRR nằm ở vị trí axit amin thứ 756 đến 759; vị trí bám của ATP GXGXXG21XK nằm từ axit amin thứ 1026 đến 1055; cluster chứa 3 tyrosine YXXYY đóng vai trò chính tạo nên vị trí phosphoryl hóa tự động nằm từ axit amin 1181 đến 1186; motif bảo thủ gắn với cơ chất của thụ thể insulin (insulin receptor substrate-1(IRS-1)-binding motif) NPEY nằm ở vị trí 997 đến 1000; vùng trình tự giàu cysteine trải dài 196 axit amin từ vị trí 179 đến 374 với 25 cystein.

Kết quả phân tích cấu trúc các protein IGFBP-1, -2, -3, -5, -6, -7 của cá tra nuôi cho thấy các IGFBP từ 1 đến 6 đều chứa domain IGFBP ở đầu N, liền ngay sau đoạn peptit tín hiệu, tiếp theo là domain nối và domain thyroglobulin tuýp 1 ở đầu C. Tuy nhiên, domain IGFBP của IGFBP-6 chỉ chứa 6 cystein (Hình 3.5), không tạo được motif “GCGCCXXC” đặc trưng như các IGFBP còn lại với 12 cystein bảo thủ (Hình 3.4, 3.5). Do các cysteine này tạo thành nhiều cầu nối bisulfit, làm ổn định cấu trúc đầu N của IGFBP [97], nên việc giảm thiểu số lượng cysteine này ở IGFBP-6 có thể là nguyên nhân làm giảm khả năng bám với IGF của protein này so với các IGFBP còn lại [98]. Trong khi ở đầu C của các IGFBP-1 đến -6 là domain thyroglobulin tuýp 1 (Hình 3.4, 3.5), thì ở IGFBP-7, tại vị trí đó lại là các domain KAZAL-FS, Ig3 và Ig (Hình 3.6). Các cysteine trong các domain này phân bố rải rác, không tạo nên motif đặc trưng ở đầu C của protein (Hình 3.6) như các IGFBP-1, -2, -3, -5, -6 (Hình 3.4, 3.5). Điều này có thể khiến cấu trúc IGFBP-7 kém bền hơn so với các protein còn lại vốn có các cysteine bảo thủ ở cả hai đầu N và C [97], dẫn tới khả năng bám với IGF-1 và IGF-2 kém hơn hẳn các IGFBP khác [43]. Mặc dù có các domain chính tương tự nhau, các IGFBP của cá tra nuôi cũng có một số đặc điểm riêng về cấu trúc. Motif Asn-Xxx-Ser/Thr (NXS/T) trên IGFBP-3 tạo điều kiện để carbohydrate bám vào asparagines và motif Arg-Gly-Asp (RGD) ở đầu C của IGFBP-2 [58] đều được tìm thấy trong cấu trúc IGFBP-3 (XP_026800381) và IGFBP-2A (XP_026801249) của cá tra nuôi. Motif bám với heparin (PKKXRP) ở

IGFBP-2 được coi là bảo thủ ở lớp thú và lớp chim, chỉ còn một phần nhỏ ở cá ngựa vằn *D. rerio* (PK-AP) [59], còn ở IGFBP-2A của cá tra nuôi là PK-AS (XP_026801249).

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

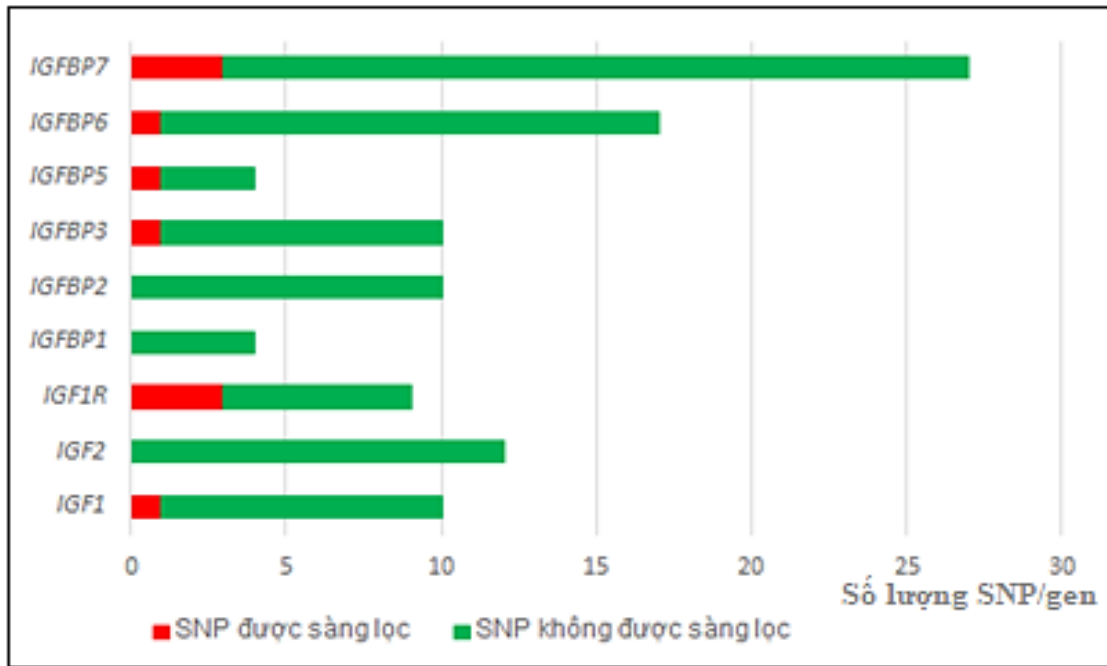
Luận án đã phát hiện SNP trên tất cả 9 gen thuộc hệ thống IGF của cá tra nuôi được phân tích, trong đó số lượng SNP trên gen *IGFBP-7* cao nhất, số lượng SNP trên gen *IGFBP-1* và *IGFBP-5* thấp nhất. Trong tổng số 103 SNP được phát hiện, có 96 SNP nằm trên các vùng không mã hóa như intron, promoter, 5'-UTR và 3'UTR (Hình 4.1, Bảng 3.1 đến 3.9), chiếm tỉ lệ khoảng 93,2%. Điều này phù hợp với thống kê từ các nghiên cứu về đa hình ở quy mô hệ gen đã khẳng định số lượng SNP nằm trong vùng không mã hóa thường cao hơn hẳn so với các SNP nằm trong vùng mã hóa, chiếm tỉ lệ khoảng 90% [99], [100], [101]. 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 *S. chuatsi* [23], gen *IGF1* trên cá chép *C. carpio* [27], gen *myostatin-1* trên cá ngỗng *A. nigrocauda* [30], gen *GH*, *IGF1* và *Myogenin (MyoG)* trên cá rô phi sông Nin *O. niloticus* cũng đưa tới nhận định này [31].



Hình 4.1. Kết quả phân tích vị trí của các SNP được phát hiện trên các gen

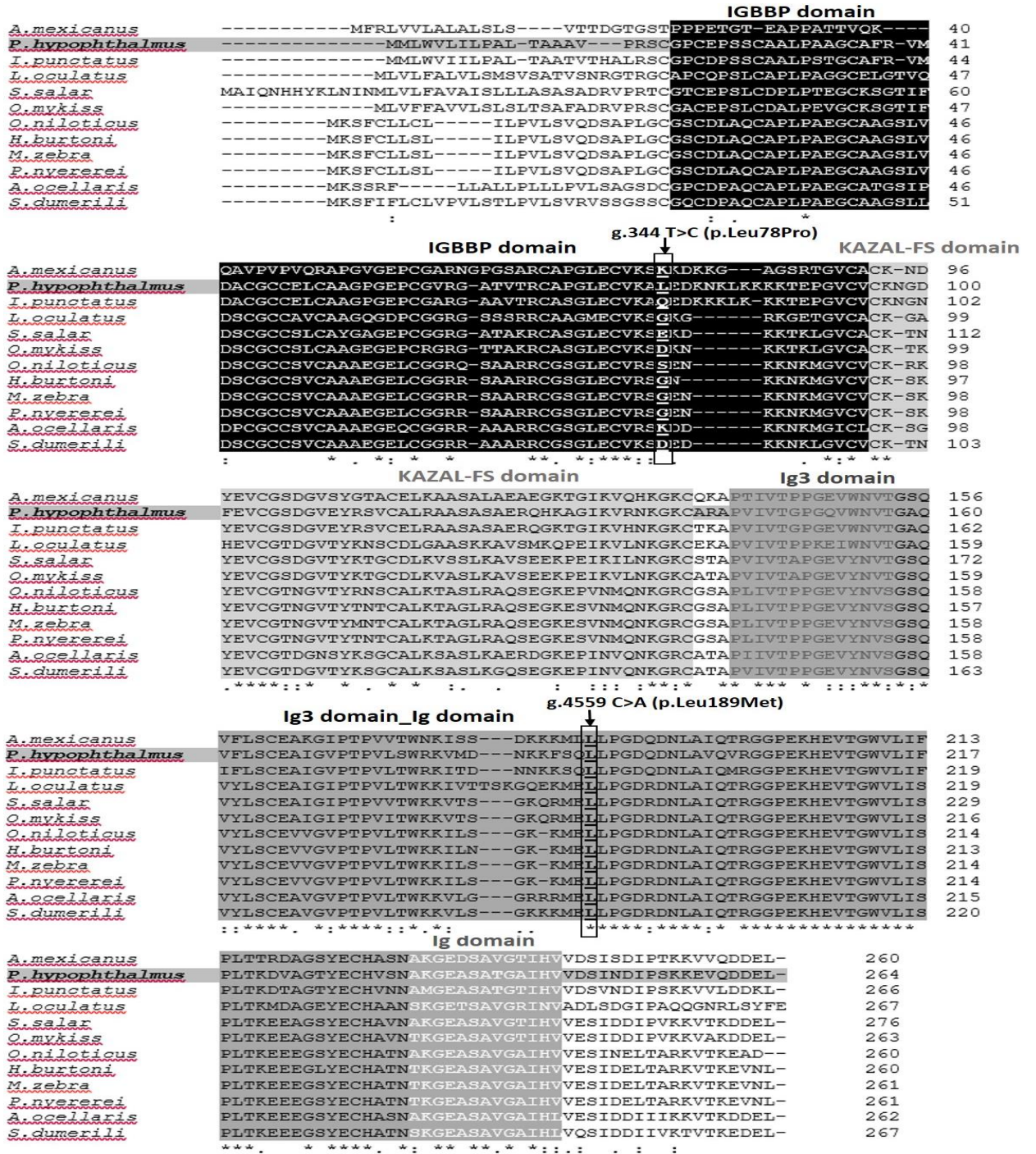
Trong số các SNP nằm trong vùng mã hóa được phát hiện trong luận án, số lượng SNP làm thay đổi trình tự axit amin và không làm thay đổi trình tự axit amin gần tương đương nhau, khác với kết quả được ở nghiên cứu trên cá hồi Đại Tây Dương *S. salar* chỉ ra số lượng SNP gây thay đổi trình tự axit amin chiếm tỉ lệ cao hơn [102], hoặc ở cá tuyết Đại Tây Dương *Gadus morhua* khi khẳng định SNP gây thay đổi trình tự axit amin chỉ chiếm 36% số lượng SNP ở vùng mã hóa [101]. Tuy vậy, số lượng SNP gây thay đổi trình tự axit amin cũng chỉ chiếm 3,8 % tổng số SNP được phát hiện trên các gen đích của cá tra nuôi (4/103 SNP), đây là một tỉ lệ nhỏ, có thể là do khi 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 [103]. 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 [104], 10% ở cá nheo Mỹ *I. punctatus* [105], cá bơn *Solea solea L.* [103], 11% đến 14% ở cá bơn Đại Tây Dương *Scophthalmus maximus* [106], [107].

Trong tổng số 103 SNP được phát hiện trên 9 gen, chỉ có 10 SNP được sàng lọc thuộc các gen *IGF1*, *IGF1R*, *IGFBP-3*, *-5*, *-6* và *IGFBP-7* (Hình 4.2), (Bảng 3.10) tiếp tục được xác định với bộ mẫu lớn hơn, chiếm tỉ lệ 9,7%. Do các đa hình xuất hiện trong quần thể chủ yếu là các biến dị mang tính chất trung tính, có thể không có tác động rõ rệt làm phân hóa trính trạng trong quần thể đó, nên 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. Điều này đã được chứng minh trong các nghiên cứu trên nhiều loài cá xương như: cá hồi Đại Tây Dương *S. salar* khi chỉ có một bộ chỉ thị 60 SNP trong tổng số 7000 SNP được sử dụng để xác định cá hồi trong tự nhiên và cá hồi nuôi [108], trên cá nheo lục *Ictalurus furcatus* khi sử dụng 64 SNP trên tổng số 4275 SNP để genotyping xác định cá tự nhiên và cá được gia hóa [109], trên cá rô mo Trung Quốc *S. chuatsi* khi sử dụng 48 trên tổng số 5205 SNP để đánh giá các tính trạng sinh trưởng như trọng lượng, chiều dài, độ dày, chiều cao thân của loài cá này [110].



Hình 4.2. Tỷ lệ các SNP được sàng lọc trên tổng số SNP được phát hiện trên các gen

Không có SNP được sàng lọc trên các gen *IGF2*, *IGFBP-1* và gen *IGFBP-2* của cá tra nuôi (Hình 4.2, các kết quả ở mục 3.3.2, 3.3.4, 3.3.5). 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 của cá xương. Cụ thể, biểu hiện của gen *IGF2* liên quan đến các tính trạng tăng trưởng của cá nheo Mỹ *I. punctatus* [74], cá lười trâu *Cynoglossus semilaevis* [75], cá rô phi sông Nin *O. niloticus* [76], cá vược châu Âu *D. labrax*, L. [77]; sự tăng cường biểu hiện của gen *IGFBP-1a* ở phôi của các loài cá thuộc họ cá chép và của gen *IGFBP-2a* và *IGFBP-2b* ở cá ngựa vằn *D. rerio* được chứng minh gây nên sự chậm tăng trưởng và phát triển [41]. Trong khi đó, có rất ít nghiên cứu chỉ ra sự mối liên quan chặt chẽ SNP trên gen *IGF2* với các tính trạng tăng trưởng của cá xương [78], [79] và chưa có nghiên cứu nào về mối tương quan giữa SNP trên gen *IGFBP-1* và *IGFBP-2* với các tính trạng tăng trưởng của cá cho kết quả khả quan. Do đó, kết quả trên thu được trong luận án khá phù hợp với tình hình nghiên cứu thế giới. Riêng đối với gen *IGF2*, có thể giải thích điều này do sự đa dạng của gen *IGF2* ở cá rất hạn chế, dẫn tới cấu trúc khá đơn giản, ít biến đổi của gen này [111].



Hình 4.3. Kết quả so sánh các trình tự protein IGFBP-7 của cá tra nuôi (*Pangasianodon hypophthalmus*) và các loài cá xương khác

Astyanax mexicanus: cá hang mù Mexico, *Ictalurus punctatus*: cá nhèo Mỹ, *Lepisosteus oculatus*: cá lạng đốm, *Salmo salar*: cá hồi Đại Tây Dương, *Oncorhynchus mykiss*: cá hồi vân, *Oreochromis niloticus*: cá rô sông Nin, *Haplochromis burtoni*: Burton's mouth-brooder (một loài cá nước ngọt), *Maylandia zebra*: zebra mbuna (một loài cá cảnh), *Pundamilia nyererei*: một loài cá nước ngọt thuộc họ Cichlidae, *Amphiprion ocellaris*: cá hề ocellaris, *Seriola dumerili*: cá cam.

Một trong những tiêu chí quan trọng để SNP được sàng lọc đó là gây thay đổi trình tự axit amin. Khi đó, các vị trí axit amin bị thay thế trên protein tương ứng sẽ được xác định có nằm trong vùng chức năng đặc biệt của protein hay không để bước đầu đánh giá mức độ ảnh hưởng của SNP lên cấu trúc, chức năng của protein. Trong luận án, 4 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*, tương ứng theo kết quả của phần mềm xác định chuỗi tín hiệu SignalP-5.0 (Phụ lục 2). Do chuỗi peptit tín hiệu đóng một vai trò quan trọng trong việc vận chuyển và bài tiết protein *IGFBP-3* [112] và *IGFBP-5* [113], nên tác động của những sự thay đổi trình tự axit amin trên lên chức năng của hai protein này sẽ cần tiếp tục được nghiên cứu sâu hơn nếu 2 SNP *IGFBP-3*. 704C>G và *IGFBP-5*. 525T>A thể hiện được sự liên quan đến tính trạng tăng trưởng của cá tra nuôi. Luận án cũng tiến hành so sánh trình tự protein *IGFBP-7* của cá tra nuôi và các trình tự *IGFBP-7* của các loài cá xương khác (Hình 4.3), qua đó nhận 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 [114](Hình 4.3).

Ngoài 4 SNP của gen *IGFBP-3*, -5, -7 gây thay đổi trình tự axit amin trên protein tương ứng, 6 SNP còn lại được sàng lọc đều nằm trong các intron của các gen (Bảng 3.10). Nếu thể hiện được sự liên quan đến tính trạng tăng trưởng của cá tra nuôi, tác động của các SNP này đến sự điều hòa mức độ biểu hiện của gen/protein tương ứng cần được nghiên cứu sâu hơn ở các thí nghiệm tiếp theo. Nghiên cứu phát hiện và genotyping các SNP trên gen *IGF1* để xác định sự liên quan đến tính trạng tăng trưởng của cá chép *C. carpio* đã phát hiện được 4 SNP, trong đó có 3 SNP thuộc intron 2 và 1 SNP gây thay đổi trình tự axit amin [27], trên cá hồi Đại Tây Dương *S. salar* phát hiện được 3 SNP trong vùng không mã hóa gồm 1 SNP thuộc vùng promoter và 2 SNP thuộc các intron [26], trên cá rô phi sông Nin *O. niloticus* phát hiện 4/8 SNP thuộc intron, 4 SNP còn lại

thuộc promoter [31]. Như vậy, kết quả phát hiện các SNP trên gen *IGF1* của cá tra nuôi đều nằm trong intron (Bảng 3.1), trong đó có 1 SNP thuộc intron 2 được sàng lọc để genotyping trong bộ mẫu lớn hơn (Bảng 3.10) là khá phù hợp với các nghiên cứu nêu trên với tỉ lệ SNP trên intron của gen *IGF1* chiếm đa số. Đối với gen *IGFIR*, so sánh với kết quả SNP genotyping trên 195 cá thể cá bản địa Trung Quốc *Odontobutis potamophila* đã phát hiện 3 SNP đều nằm trong vùng intron [80], luận án cũng sàng lọc được 2 SNP nằm trong intron 1 và 1 SNP nằm trong intron 15 của gen này ở cá tra nuôi để kiểm nghiệm sự liên quan đến tính trạng tăng trưởng (Bảng 3.10). Mặc dù nghiên cứu tìm đa hình trên gen *IGFBP-6* liên quan đến tăng trưởng của cá xương chưa từng được tiến hành, tuy nhiên Fang và cộng sự (2015b) đã khuếch đại các đoạn chứa 257 bp của intron 3, exon 4 và 170 bp của đoạn 3'UTR của gen này trên các giống lợn nhỏ Bama và Tibetan (Trung Quốc), lợn Landrace, lợn Yorkshire và lợn rừng Đông Bắc và phát hiện ra 3 trên 6 SNP thuộc ở intron 3, sau đó kiểm nghiệm mối tương quan của các SNP này lên tính trạng tăng trưởng của các giống lợn trên [115]. Kết quả thu được trong luận án tương đồng khi cũng sàng lọc được SNP trên intron 3 của gen *IGFBP-6* để tiếp tục kiểm nghiệm sự liên quan đến tính trạng tăng trưởng của cá tra nuôi (Bảng 3.10).

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

Sáu chỉ thị SNP tiềm năng trên các gen thuộc hệ thống IGF liên quan đến tính trạng tăng trưởng của cá tra nuôi được so sánh với các SNP liên quan đến tính trạng tăng trưởng của các loài cá xương khác trong các nghiên cứu trên thế giới. Cụ thể, SNP 13680 A>T ở intron 2 của gen *IGF1* thể hiện sự liên quan đến tính trạng tăng trưởng của cá tra nuôi. Điều này 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 không chỉ các loài cá xương như cá chép *C. carpio* [27], cá hồi Đại Tây Dương *S. salar* [26], mà còn trên các loài gia súc [116], [117] và chim sẻ [118].

Trong số 3 SNP được sàng lọc và genotyping của gen *IGFIR* 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 nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm. 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 1208G>A liên quan đến các tính trạng tăng trưởng quan trọng như chiều dài và khối lượng cơ thể, chiều rộng thân với giá trị p nhỏ hơn 0,05 [80].

Phần lớn các SNP trên gen *IGFBP-3* và *IGFBP-5* liên quan đến tính trạng tăng trưởng đều nằm trong vùng không mã hóa, chủ yếu là promoter, ví dụ như ở nghiên cứu trên người [119], lợn [120], cá chép [121], gà [122], hoặc ở trên intron trong nghiên cứu trên lợn [123]. Như vậy, kết quả của luận án đã có sự khác biệt khi tìm được 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 gây sự thay thế axit amin với tính trạng tăng trưởng của các loài thủy sản. Ví dụ, tất cả 3 đa hình gây sự thay thế axit amin c.543C > T (Phe181Leu), c.848A > G (Tyr283Cys) và c.1625G > C (Gly542Phe) trên gen *FOXO* đều thể hiện mối liên quan tới các tính trạng khối lượng cơ thể, chiều dài, độ rộng và chiều cao vò của ngao [124]. Đối với các cá thể con lai của *Culter alburnus* (♀) x *Ancherythroculter nigrocauda* (♂), các phép phân tích đã chỉ ra SNP (c.6T>C) trên exon 2 của gen *MTTN*, gây sự biến đổi axit amin Serine thành Proline, có mối liên quan chặt chẽ với tính trạng chiều dài tổng, chiều dài thân, chiều cao thân, chiều dài đầu và khối lượng cơ thể [125]. Một trong hai SNP gây sự thay thế axit amin được sàng lọc trên gen *Myogenin* là *MYOG131* (g.529T>A) được coi là có mối liên quan đến sự tăng cường tăng trưởng của cá nheo Mỹ [105].

Từ dữ liệu SNP của bộ mẫu khởi tạo, có một SNP 2278 C>A ở trên gen *IGFBP-6* được sàng lọc nhưng SNP này lại không thể hiện sự khác biệt về thành phần kiểu gen hay thành phần alen giữa nhóm tăng trưởng nhanh và nhóm tăng trưởng chậm của cá tra nuôi khi được phân tích trên 160 cá thể. 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 [41].

Trong số 3 SNP được sàng lọc của gen *IGFBP-7*, SNP 2060 A>G trên intron 2 và SNP 4559 C>A (p.Leu189Met) trên vùng mã hóa làm thay đổi trình tự axit amin thể hiện được mối liên quan đến tính trạng tăng trưởng của cá tra nuôi. Đây là một kết quả phù

hợp với nhiều nghiên cứu khác trên thế giới khi chỉ ra được mối liên quan chặt chẽ giữa tính trạng tăng trưởng của cá xương với các SNP trên intron [103], [23], [27], [126], hay với các SNP làm thay đổi trình tự axit amin [127], [128].

Các giá trị PIC, MAF của 6 chỉ thị SNP tiềm năng được tính toán trên tổng số 160 cá thể cho thấy các giá trị PIC của các SNP đa phần nằm trong khoảng 0,25 đến 0,5 cho thấy mức độ đa dạng di truyền trung bình [87], là cơ sở tiềm năng cho việc sử dụng chúng trong chọn giống. Phần lớn các giá trị MAF của các SNP cao, 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 [88]. 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 (Bảng 3.18), nguyên nhân có thể do SNP này gây thay đổi trình tự axit amin trong vùng bảo thủ của domain Ig-Ig3 của protein *IGFBP-7* (Hình 4.3, mục 4.2.2) [129].

Từ mỗi gen *IGF1R* và *IGFBP-7* của cá tra nuôi, luận án đều sàng lọc được 3 SNP để đánh giá sự mất cân bằng liên kết hay khả năng kết hợp không ngẫu nhiên giữa các SNP được sàng lọc trên cùng một gen này. Sự mất cân bằng liên kết giữa hai alen thuộc các locus khác nhau có thể liên quan đến các sự kiện đột biến, sự tái tổ hợp, khoảng cách di truyền, chọn lọc trên quần thể và lịch sử phát triển của quần thể đó [89], [130]. Kết quả cho thấy các SNP trên cùng một gen đều có thể kết hợp không ngẫu nhiên với nhau và được di truyền cùng nhau, từ đó củng cố cho các phân tích về các kiểu gen đơn bội tạo nên từ các SNP đó.

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 (Bảng 3.12, mục 3.4.2.2.a), 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 (Bảng 3.14, mục 3.4.2.2). 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 (Bảng 3.18, mục 3.4.2.6.a), 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 cho mỗi nhóm tăng trưởng nhanh/chậm (Bảng 3.20, mục 3.4.2.6.c). 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 [131], [132] và loài khác [133], hay với các tính trạng kinh tế quan trọng [134]. Ngoài ra, nghiên cứu trên các gen thuộc hệ thống IGF trên các loài cá cũng đã chỉ ra haplotype tạo nên từ các SNP trên cùng một gen *IGF1* của cá vược biển đen *Micropterus salmoides* liên quan chặt chẽ đến khối lượng cơ thể [135], haplotype phát sinh từ các SNP trên gen *IGF1* của cá hồi Đại Tây Dương *S. salar* ảnh hưởng đến cân nặng thu hoạch của loài cá này [26].

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

Các haplotype, tổ hợp kiểu gen cung cấp thông tin rất ý nghĩa để tìm mối tương quan giữa các SNP trên gen đích và tính trạng tăng trưởng của cá [135], [26], [27], [136]. Trong nghiên cứu này, tác động tổng hợp của 6 SNP tiềm năng liên quan đến tính trạng tăng trưởng đã được phân tích thông qua sự khác biệt về tỉ lệ haplotype cũng như tổ hợp kiểu gen giữa hai nhóm tăng trưởng nhanh và tăng trưởng chậm. Kết quả đã chỉ ra các haplotype chiếm tỉ lệ cao ở nhóm tăng trưởng nhanh/chậm. Các giá trị p đánh giá sự khác biệt về tần số các haplotype tạo nên từ 6 SNP tiềm năng giữa hai nhóm tăng trưởng nhanh/chậm phần lớn đều thể hiện sự khác biệt có ý nghĩa (11/14 haplotype có giá trị p nhỏ hơn 0,05) (Bảng 3.21). Điều này cho thấy việc phân tích các haplotype tạo nên từ các chỉ thị SNP tiềm năng vẫn có thể đem lại hiệu quả tốt, nhằm bổ sung thêm các chỉ thị phân tử liên quan đến tính trạng quan tâm.

Tổ hợp kiểu gen *IGF1*. 13680 AT, *IGF1R*. 13357 CT, *IGFBP-3*. 704 GC, *IGFBP-5*. 525 AT, *IGFBP-7*. 2060 GA và *IGFBP-7*. 4559 AA (tạo nên từ haplotype CO_H2 và CO_H11 (diplotype H2/H11)) có tần số cao hơn ở nhóm tăng trưởng nhanh và tổ hợp kiểu gen *IGF1*. 13680 AT, *IGF1R*. 13357 TT, *IGFBP-3*. 704 GC, *IGFBP-5*. 525 AT, *IGFBP-7*. 2060 GG và *IGFBP-7*. 4559 AA (tạo nên từ haplotype CO_H2 và CO_H5 (diplotype H2/H5)) chỉ xuất hiện ở nhóm tăng trưởng nhanh (Bảng 3.22). Kết quả này phù hợp với kết quả về thành phần kiểu gen liên quan đến tính trạng tăng trưởng của mỗi chỉ thị SNP trên 160 cá thể (các kết quả ở mục 3.4.2.1, 3.4.2.2.a, 3.4.2.3, 3.4.2.4 và 3.4.2.6.a) và kết quả về tần số haplotype tạo nên từ 6 chỉ thị SNP (Bảng 3.21, mục

3.4.2.7.a), cho thấy tính thống nhất của các kết quả nghiên cứu được trình bày trong luận án.

Tác động tổng hợp của các SNP trên nhiều gen khác nhau thuộc hệ thống IGF lên tính trạng tăng trưởng của cá tra nuôi được nêu trên đã góp phần chỉ ra tính trạng tăng trưởng của cá tra nuôi là tính trạng đa gen, tương tự như tính trạng tăng trưởng của các loài cá khác như cá hồi vân *O. mykiss* [137], cá chép *C. carpio* [138]. Như vậy, việc phát triển chỉ thị phân tử dạng panel, có tích hợp nhiều SNP tiềm năng trên các gen liên quan khác nhau để đánh giá tính trạng tăng trưởng, phục vụ công tác chọn giống cá tra nuôi là có cơ sở khoa học và rất ý nghĩa, cần được tiến hành song song với việc tìm kiếm các chỉ thị SNP riêng lẻ.

KẾT LUẬN

1. Đã phân tích được trình tự và cấu trúc của 9 gen thuộc hệ thống IGF (bao gồm *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6, -7, xác định và giải mã được các vùng gen chứa SNP liên quan đến tính trạng tăng trưởng.
2. Đã phân tích và phát hiện được 103 SNP trên 9 gen (*IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, -2, -3, -5, -6 và -7) của các cá thể cá tra thuộc bộ mẫu khởi tạo gồm cá thể tăng trưởng nhanh và cá thể tăng trưởng chậm, từ đó lựa chọn được 10/103 SNP để kiểm nghiệm trên bộ mẫu kiểm nghiệm.
3. Đã kiểm nghiệm được 10 SNP trên bộ mẫu kiểm nghiệm bằng phương pháp Kéo dài một nucleotide (Single base extension) thu được 6 SNP 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. Đã phân tích được 11 haplotype, các tổ hợp kiểu gen phát sinh từ các SNP có tiềm năng liên quan đến tính trạng tăng trưởng của cá tra nuôi.

KIẾN NGHỊ

Từ những kết quả đạt được, luận án đưa ra kiến nghị như sau: 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.

DANH MỤC CÔNG TRÌNH CỦA TÁC GIẢ

Tên bài báo 01: 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

Tác giả: Lê Thị Nguyên Bình, Nguyễn Thị Hoa, **Trần Thị Huyền Trang**, Nguyễn Thành Phương, Kim Thị Phương Oanh.

Tên bài báo 02: 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

Tác giả: **Trần Thị Huyền Trang**, Lê Thị Nguyên Bình, Nguyễn Thị Hoa, Trần Sơn Hoàng, Kim Thị Phương Oanh.

Tên bài báo 03: Characterization of single nucleotide polymorphism in *IGF1* and *IGFIR* genes associated with growth traits in striped catfish (*Pangasianodon hypophthalmus* Sauvage, 1878)

Tạp chí: Aquaculture 2021, 538: e736542. (doi: 10.1016/j.aquaculture.2021.736542. Epub 2021 May 30)

Tác giả: **Trang Thi Huyen Tran**, Hoa Thi Nguyen, Binh Thi Nguyen Le, Phuc Huu Tran, Sang Van Nguyen, Oanh Thi Phuong Kim.

Tên bài báo 04: Significant association between a non-synonymous SNP in *IGFBP5* gene and the growth of striped catfish (*Pangasianodon hypophthalmus*, Sauvage, 1878)

Tạp chí: Vietnam Journal of Biotechnology 2023, 21(1): 1-11 (doi: 10.15625/1811-4989/17647)

Tác giả: **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim.

Tên bài báo 05: 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)

Tạp chí: Molecular Genetics and Genomics 2023, 298(4): 883-893.

(doi:10.1007/s00438-023-02016-2. Epub 2023 Apr 25)

Tác giả: **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim.

Tên bài báo 06: Non-synonymous polymorphism in *igfbp-3* gene associated with growth traits in striped catfish (*Pangasianodon hypophthalmus*, Sauvage, 1878)

Tạp chí: Vietnam Journal of Biotechnology, accepted.

Tác giả: **Trang Thi Huyen Tran**, Binh Thi Nguyen Le, Sang Van Nguyen, Oanh Thi Phuong Kim

TÀI LIỆU THAM KHẢO

1. T.R. Roberts, V. Chavalit, *Systematic revision of the Asian catfish family Pangasiidae, with biological observations and descriptions of three new species*, Proceedings of the Academy of Natural Sciences of Philadelphia, 1991, 143, 97-143.
2. ©FAO. *Pangasianodon hypophthalmus. Cultured Aquatic Species Information Programme*, 2010
https://www.fao.org/fishery/en/culturedspecies/Pangasius_hypophthalmus/en.
3. Hiệp hội Chế biến và Xuất khẩu Thủy sản Việt Nam (VASEP). *Xuất khẩu cá tra Việt Nam năm 2021 với nhiều cố gắng ngoài sức tưởng tượng*. 2022
<https://vasep.com.vn/san-pham-xuat-khau/ca-tra/xuat-nhap-khau/xuat-khau-ca-tra-viet-nam-nam-2021-voi-nhieu-co-gang-ngoai-suc-tuong-tuong-23704>.
4. *Xuất khẩu cá tra hướng tới mục tiêu trên 1,6 tỉ USD trong năm 2022*. 2022
<https://www.mard.gov.vn/Pages/xuat-khau-ca-tra-huong-toi-muc-tieu-tren-1-6-ti-usd-trong-nam-2022.aspx>.
5. *Nhiều tín hiệu tích cực trong hoạt động chế biến và xuất khẩu cá tra*. 2022
<https://www.vietnamplus.vn/nhieu-tin-hieu-tich-cuc-trong-hoat-dong-che-bien-va-xuat-khau-ca-tra/792402.vnp>.
6. Phạm Anh Tuấn, Nguyễn Hữu Ninh, *Microsatellite nghiên cứu biến dị về màu sắc thịt trắng và thịt vàng của cá tra (Pangasius hypophthalmus)*, Hội nghị khoa học toàn quốc lần thứ 2. Tuyển tập báo cáo khoa học về nuôi trồng thủy sản, Nxb Nông nghiệp, 2003, Hà Nội.
7. Nguyễn Minh Thành, Võ Thị Minh Thư, H. Jung, et al., *Phân tích hệ gen chức năng từ mô thận cá tra (Pangasianodon hypophthalmus) nuôi ở điều kiện mặn: Lắp ráp, chú giải, phân tích chỉ thị SNP*, Tạp chí Sinh học, 2015, 37, 220-227.
8. Đào Thị Tuyết, Quyền Đình Thi, Nguyễn Thị Bảy, et al., *Đánh giá tính đa hình các quần đàn cá tra nuôi (Pangasius hypophthalmus) ở Việt nam bằng phương pháp RAPD*, Hội nghị Công nghệ Sinh học Toàn quốc, Nxb Khoa học và Kỹ thuật, 2003, Hà Nội.
9. Nguyễn Văn Cường, Nguyễn Thu Thúy, Nguyễn Thị Diệu Thúy, et al., *Phân tích di truyền AFLP cá tra (Pangasius hypophthalmus)*, Hội nghị khoa học toàn quốc lần thứ 2. Tuyển tập báo cáo khoa học về nuôi trồng thủy sản, Nxb Nông nghiệp, 2003, Hà Nội.

10. Bùi Thị Liên Hà, Nguyễn Văn Sáng, Nguyễn Văn Hào, et al., *Bước đầu đánh giá đa dạng di truyền quần đàn cá tra (Pangasianodon hypophthalmus) có nguồn gốc tự nhiên, từ các trại giống và từ chương trình chọn giống của Việt Nam*, NXB Nông nghiệp, 2011, Hà Nội.
11. S.V. Nguyen, G. Klemetsdal, J. Ødegård, et al., *Genetic parameters of filet yield and body traits on river catfish (Pangasianodon hypophthalmus) in Vietnam*, *Aquaculture*, 2007, 272, 295.
12. Nguyễn Văn Sáng, Nguyễn Văn Hào, Trần Đình Trọng, et al., *Chọn giống cá tra (Pangasianodon hypophthalmus) nhằm tăng tỷ lệ phi lê bằng chọn lọc gia đình*, Báo cáo khoa học Viện Nghiên cứu Nuôi trồng Thủy sản 2, 2009, Thành phố Hồ Chí Minh.
13. Nguyễn Văn Sáng, Nguyễn Văn Hào, Nguyễn Quyết Tâm, et al., *Chọn giống cá tra theo tính trạng khối lượng thu hoạch*, Tạp chí Nông nghiệp và Phát triển Nông thôn, 2015, 6, 147-156.
14. N.T. Vu, N. Van Sang, T.H. Phuc, et al., *Genetic evaluation of a 15-year selection program for high growth in striped catfish Pangasianodon hypophthalmus*, *Aquaculture*, 2019, 509, 221–226.
15. Nguyễn Thanh Vũ, Trịnh Quốc Trọng, Trần Hữu Phúc, et al., *Ứng dụng di truyền phân tử và di truyền số lượng chọn giống nâng cao tốc độ sinh trưởng cá Tra (Pangasianodon hypophthalmus)*, Kỷ yếu Hội nghị Khoa học trẻ Thủy sản Toàn quốc lần thứ VIII – Youth Fish 2017, 2017, 109-119.
16. Kim Thị Phương Oanh, Nguyễn Thành Phương, E. Shoguchi, et al., *A draft genome of the striped catfish, Pangasianodon hypophthalmus, for comparative analysis of genes relevant to development and a resource for aquaculture improvement*, *BMC Genomics*, 2018, 19, 733.
17. Kim Thị Phương Oanh, *Phân tích hệ gen biểu hiện (exome + transcriptome) của cá tra nhằm phát triển chỉ thị phân tử phục vụ chọn giống cá tra theo hướng tăng trưởng*, Viện Nghiên cứu hệ gen-Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 2018, Hà Nội.

18. Kim Thị Phương Oanh, *Phân tích hệ gen của cá tra nhằm phát triển chỉ thị phân tử phục vụ chọn giống cá tra theo hướng tăng trưởng (giai đoạn 2)*, Viện Nghiên cứu hệ gen-Viện Hàn lâm Khoa học và Công nghệ Việt Nam, 2020, Hà Nội.
19. H. Marnis, *Correlation of microsatellite DNA markers with growth traits in striped catfish (Pangasianodon hypophthalmus)*, Indonesian Aquaculture Journal, 2018, 13, 51-56.
20. I.A. Johnston, N.I. Bower, D.J. Macqueen, *Growth and the regulation of myotomal muscle mass in teleost fish*, The Journal of Experimental Biology, 2011, 214, 1617-1628.
21. E.J. Vélez, E. Lutfi, S. Azizi, et al., *Understanding fish muscle growth regulation to optimize aquaculture production*, Aquaculture, 2017, 467, 28-40.
22. C. De-Santis, D.R. Jerry, *Candidate growth genes in finfish — Where should we be looking?*, Aquaculture, 2007, 272, 22-38.
23. C. Tian, M. Yang, L. Lv, et al., *Single nucleotide polymorphisms in growth hormone gene and their association with growth traits in Siniperca chuatsi (Basilewsky)*, International Journal of Molecular Sciences, 2014, 15, 7029-7036.
24. C.F. Sun, H.L. Sun, J.J. Dong, et al., *Correlation analysis of mandarin fish (Siniperca chuatsi) growth hormone gene polymorphisms and growth traits*, Journal of Genetics, 2019, 98, 58.
25. R. Wenne, *Single nucleotide polymorphism markers with applications in aquaculture and assessment of its impact on natural populations*, Aquatic Living Resources, 2018, 31, 2.
26. H.Y. Tsai, A. Hamilton, D.R. Guy, et al., *Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF1) gene are associated with growth-related traits in farmed Atlantic salmon*, Animal Genetics, 2014, 45, 709-715.
27. X. Feng, X. Yu, J. Tong, *Novel single nucleotide polymorphisms of the insulin-like growth factor-I gene and their associations with growth traits in common carp (Cyprinus carpio L.)*, International Journal of Molecular Sciences, 2014, 15, 22471-22482.
28. E. Özcan Gökçek, R. Işık, *Associations between genetic variants of the insulin-like growth factor I (IGF-I) gene and growth traits in European sea bass (Dicentrarchus labrax, L.)*, Fish Physiology and Biochemistry, 2020, 46, 1131-1138.

29. Y. Sun, X. Yu, J. Tong, *Polymorphisms in Myostatin Gene and associations with growth traits in the common carp (Cyprinus carpio L.)*, International Journal of Molecular Sciences, 2012, 13, 14956-14961.
30. Y. Sun, Q. Li, G. Wang, et al., *Polymorphisms in the Myostatin-1 gene and their association with growth traits in Ancherythroculter nigrocauda*, Chinese Journal of Oceanology and Limnology, 2017, 35, 597-602.
31. B.L. Cuevas-Rodríguez, A.M. Sifuentes-Rincón, P. Ambriz-Morales, et al., *Novel single nucleotide polymorphisms in candidate genes for growth in tilapia (Oreochromis niloticus)*, Revista Brasileira de Zootecnia, 2016, 45, 345-348.
32. A.W. Wood, C. Duan, H.A. Bern, *Insulin-like growth factor signaling in fish*, International review of cytology, 2005, 243, 215-285.
33. D. LeRoith, C. Bondy, S. Yakar, et al., *The somatemedin hypothesis: 2001*, Endocrine Reviews, 2001, 22, 53-74.
34. N. Bower, X. Li, R. Taylor, et al., *Switching to fast growth: the insulin-like growth factor (IGF) system in skeletal muscle of Atlantic salmon*, The Journal of Experimental Biology 2008, 211, 3859-3870.
35. E.N. Fuentes, J.A. Valdés, A. Molina, et al., *Regulation of skeletal muscle growth in fish by the growth hormone – Insulin-like growth factor system*, General and Comparative Endocrinology, 2013, 192, 136-148.
36. N.I. Bower, I.A. Johnston, *Transcriptional regulation of the IGF signaling pathway by amino acids and insulin-like growth factors during myogenesis in Atlantic salmon*, PLoS One, 2010, 5, e11100.
37. C.H. Chu, B.S. Tzang, L.M. Chen, et al., *IGF-II/mannose-6-phosphate receptor signaling induced cell hypertrophy and atrial natriuretic peptide/BNP expression via G alpha q interaction and protein kinase Calpha/CaMKII activation in H9c2 cardiomyoblast cells*, Journal of Endocrinology, 2008, 197, 381-390.
38. C. Duan, *Specifying the cellular responses to IGF signals: roles of IGF-binding proteins*, Journal of Endocrinology, 2002, 175, 41-54.
39. S.M. Firth, R.C. Baxter, *Cellular actions of the insulin-like growth factor binding proteins*, Endocrine Reviews, 2002, 23, 824-854.

40. S. Kajimura, K. Aida, C.M. Duan, *Insulin-like growth factor-binding protein-1 (IGFBP-1) mediates hypoxia-induced embryonic growth and developmental retardation.*, Proceedings of the National Academy of Sciences of the United States of America 2005, 102, 1240-1245.
41. D.G. de la Serrana, D.G. Macqueen, *Insulin-Like Growth Factor-Binding Proteins of Teleost Fishes*, Frontiers in Endocrinology 2018, 9, 80.
42. M.R. Schneider, E. Wolf, A. Hoefflich, et al., *IGF-binding protein-5: flexible player in the IGF system and effector on its own*, Journal of Endocrinology, 2002, 172, 423-440.
43. D.V. Nguyen, S. Li Calzi, L.C. Shaw, et al., *An ocular view of the IGF-IGFBP system*, Growth Hormone & IGF Research, 2013, 23, 45-52.
44. Y. Yamanaka, E.M. Wilson, R.G. Rosenfeld, et al., *Inhibition of insulin receptor activation by insulin-like growth factor binding proteins*, Journal of Biological Chemistry, 1997, 272, 30729-30734.
45. H.E. Lau, S.H. Chalasani, *Divergent and convergent roles for insulin-like peptides in the worm, fly and mammalian nervous systems*, Invertebrate Neuroscience, 2014, 14, 71-78.
46. W. Ruan, E. Xu, F. Xu, et al., *IGFBP7 plays a potential tumor suppressor role in colorectal carcinogenesis*, Cancer Biology & Therapy, 2007, 6, 354-359.
47. T. Benatar, W. Yang, Y. Amemiya, et al., *IGFBP7 reduces breast tumor growth by induction of senescence and apoptosis pathways*, Breast cancer research and treatment, 2011, 133, 563-573.
48. C. Yue, M. Yang, Q. Tian, et al., *IGFBP7 is associated to prognosis and could suppress cell survival in cholangiocarcinoma*, Artificial Cells, Nanomedicine, and Biotechnology, 2018, 46, 817-825.
49. M.G. Vizioli, M. Sensi, C. Miranda, et al., *IGFBP7: an oncosuppressor gene in thyroid carcinogenesis*, Oncogene, 2010, 29, 3835-3844.
50. L. Sullivan, T.M. Murphy, C. Barrett, et al., *IGFBP7 promoter methylation and gene expression analysis in prostate cancer*, The Journal of Urology, 2012, 188, 1354-1360.

51. M. Akiel, C. Guo, X. Li, et al., *IGFBP7 Deletion Promotes Hepatocellular Carcinoma*, *Cancer Research*, 2017, 77, 4014-4025.
52. J. Kim, W.H. Kim, S.J. Byeon, et al., *Epigenetic Downregulation and Growth Inhibition of IGFBP7 in Gastric Cancer*, *Asian Pacific Journal of Cancer Prevention*, 2018, 19, 667-675.
53. Z. Lu, J. Chiu, L.R. Lee, et al., *Reprogramming of human fibroblasts into osteoblasts by insulin-like growth factor-binding protein 7*, *Stem Cells Translational Medicine*, 2020, 9, 403-415.
54. J. Li, J. Liu, G. Campanile, et al., *Novel insights into the genetic basis of buffalo reproductive performance*, *BMC Genomics*, 2018, 19, 814.
55. B.B. Kamangar, J.C. Gabillard, J. Bobe, *Insulin-like growth factor-binding protein (IGFBP)-1, -2, -3, -4, -5, and -6 and IGFBP-related protein 1 during rainbow trout postvitellogenesis and oocyte maturation: molecular characterization, expression profiles, and hormonal regulation*, *Endocrinology*, 2006, 147, 2399-2410.
56. A. Alzaid, S.A.M. Martin, D.J. Macqueen, *The complete salmonid IGF-IR gene repertoire and its transcriptional response to disease*, *Scientific Reports*, 2016, 6, 34806.
57. T. Maures, S.J. Chan, B. Xu, et al., *Structural, biochemical, and expression analysis of two distinct insulin-like growth factor I receptors and their ligands in zebrafish*, *Endocrinology*, 2002, 143, 1858 – 1871.
58. D.O. Daza, G. Sundström, C.A. Bergqvist, et al., *Evolution of the Insulin-like Growth Factor Binding Protein (IGFBP) family*, *Endocrinology*, 2011, 152, 2278-2289.
59. T.J. Maures, C. Duan, *Structure, Developmental Expression, and Physiological Regulation of Zebrafish IGF Binding Protein-1*, *Endocrinology*, 2002, 143, 2722-2731.
60. P.A. Morin, G. Luikart, R.K. Wayne, et al., *SNPs in ecology, evolution and conservation*, *Trends in Ecology & Evolution*, 2004, 19, 208-216.
61. Z.J. Liu, J.F. Cordes, *DNA marker technologies and their applications in aquaculture genetics*, *Aquaculture*, 2004, 238, 1-37.
62. K.D. Rasal, V. Chakrapani, A.K. Pandey, et al., *Status and future perspectives of single nucleotide polymorphisms (SNPs) markers in farmed fishes: Way ahead using next generation sequencing*, *Gene Reports*, 2017, 6, 81-86.

63. Z.J. Liu, *Next Generation Sequencing and Whole Genome Selection in Aquaculture*, Wiley-Blackwell,, 2011, New York.
64. I. Okumus, Y. Ciftci, *Fish population genetics and molecular markers: II. Molecular markers and their applications in fisheries and aquaculture*, Journal of Fisheries and Aquatic Science, 2003, 3, 51-79.
65. M. Hakim, S. Ahmad, *Advances in Molecular Markers and Their Applications in Aquaculture and Fisheries*, Genetics of Aquatic Organisms, 2017, 1, 27-41.
66. P.Y. Kwok, *Methods for genotyping single nucleotide polymorphisms*, Annual Review of Genomics and Human Genetics, 2001, 2, 235-258.
67. R. Kalendar, A. Baidyussen, D. Serikbay, et al., *Modified "Allele-Specific qPCR" Method for SNP Genotyping Based on FRET*, Frontiers in Plant Science, 2022, 12, 747886.
68. A.C. Syvanen, *From gels to chips: "minisequencing" primer extension for analysis of point mutations and single nucleotide polymorphisms*, Human Mutation, 1999, 13, 1-10.
69. C. Phillips, A. Salas, J.J. Sánchez, et al., *Inferring ancestral origin using a single multiplex assay of ancestry-informative marker SNPs*, Forensic Science International: Genetics, 2007, 1, 273-280.
70. F. Fiorentino, *The minisequencing method: an alternative strategy for preimplantation genetic diagnosis of single gene disorders*, Molecular Human Reproduction, 2003, 9, 399-410.
71. T.M. Nelson, R.S. Just, O. Loreille, et al., *Development of a multiplex single base extension assay for mitochondrial DNA haplogroup typing*, Croatian medical journal, 2007, 48, 460-472.
72. P.M. Vallone ,J.M. Butler, *Y-SNP typing of U.S. African American and Caucasian samples using allele-specific hybridization and primer extension*, Journal of Forensic Sciences, 2004, 49, 723-732.
73. J. Di Cristofaro, M. Silvy, J. Chiaroni, et al., *Single PCR multiplex SNaPshot reaction for detection of eleven blood group nucleotide polymorphisms: optimization, validation,*

and one year of routine clinical use, The Journal of Molecular Diagnostics, 2010, 12, 453-460.

74. B.C. Peterson, G.C. Waldbieser, L. Bilodeau, *IGF-I and IGF-II mRNA expression in slow and fast growing families of USDA103 channel catfish (Ictalurus punctatus)*, Comparative Biochemistry and Physiology - Part A: Molecular & Integrative Physiology, 2004, 139, 317-323.

75. Q. Ma, S.F. Liu, Z.M. Zhuang, et al., *Genomic structure, polymorphism and expression analysis of growth hormone-releasing hormone and pituitary adenylate cyclase activating polypeptide genes in the half-smooth tongue sole (Cynoglossus semilaevis)*, Genetics and Molecular Research, 2011, 10, 3828-3846.

76. C. Huang, Y.-H. Li, S.-Y. Hu, et al., *Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia*, Journal of Animal Science, 2012, 90, 4266–4279.

77. B. Louro, R.S.T. Martins, P.I.S. Pinto, et al., *SuperSAGE digital expression analysis of differential growth rate in a European sea bass population*, Aquaculture and Fisheries, 2019, 4, 17-26.

78. Y. Juhua, C. Xuefeng, L. Jianlin, et al., *Isolation of IGF2 and association of IGF2 polymorphism with growth trait in genetically improved farmed tilapias, Oreochromis niloticus L*, Aquaculture Research, 2010, 41, 743-750.

79. T. Teng, X. Zhao, C. Li, et al., *Cloning and expression of IGF-I, IGF-II, and GHR genes and the role of their single-nucleotide polymorphisms in the growth of pikeperch (Sander lucioperca)*, Aquaculture International, 2020, 28, 1547-1561.

80. H. Zhang, C. Zhao, S. Yin, et al., *Characterization and Identification of Single Nucleotide Polymorphism Within the IGF-1R Gene Associated with Growth Traits of Odontobutis potamophila*, Journal of the World Aquaculture Society, 2018, 49, 366-379.

81. A.J.J. Almagro, K.D. Tsirigos, C.K. Sønderby, et al., *SignalP 5.0 improves signal peptide predictions using deep neural networks*, Nature Biotechnology, 2019, 37, 420-423.

82. *Phenolic DNA Purification – Background and Protocol*.
<https://www.carlroth.com/medias/Info-Brochure-PhenolicPurification-EN.pdf>.

83. A. Untergasser, I. Cutcutache, T. Koressaar, et al., *Primer3--new capabilities and interfaces*, Nucleic Acids Research, 2012, 40, e115.
84. T.A. Hall, *BioEdit: A user-friendly biological sequence alignment editor and analysis program for Windows 95/98/NT*, Nucleic Acids Symposium Series, 1999, 41, 95-98.
85. R.C. Edgar, *MUSCLE: multiple sequence alignment with high accuracy and high throughput*, Nucleic Acids Research, 2004, 32, 1792-1797.
86. Y.Y. Shi, L. He, *SHEsis, a powerful software platform for analyses of linkage disequilibrium, haplotype construction, and genetic association at polymorphism loci*, Cell Research, 2005, 15, 97-98.
87. Y. Chesnokov, A. Artem'eva, *Evaluation of the measure of polymorphism Sel'skokhozyaistvennaya Biologiya [Agricultural Biology]*, 2015, 50, 571-578.
88. J.-H. Park, M. Gail, C. Weinberg, et al., *Distribution of allele frequencies and ESs and their interrelationships for common genetic susceptibility variants*, Proceedings of the National Academy of Sciences of the United States of America, 2011, 108, 18026-18031.
89. M. Slatkin, *Linkage disequilibrium--understanding the evolutionary past and mapping the medical future*, Nature reviews. Genetics, 2008, 9, 477-485.
90. J.M. Kavran, J.M. McCabe, P.O. Byrne, et al., *How IGF-1 activates its receptor*, eLife, 2014, 3, e03772.
91. T.F. Beck, J.C. Mullikin ,L.G. Biesecker, *Systematic Evaluation of Sanger Validation of Next-Generation Sequencing Variants*, Clinical Chemistry, 2016, 62, 647-654.
92. Z. Liu, S. Liu, J. Yao, et al., *The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts*, Nature Communications, 2016, 7, 11757.
93. I. Irmawati, A. Alimuddin, Z. Muhammad, et al., *Molecular cloning and sequence analysis of insulin-like growth factors-1 cDNA of giant gourami, Osphronemus goramy*, Aquaculture, Aquarium, Conservation & Legislation-International Journal of the Bioflux Society (AACL Bioflux), 2016, 9, 174-183.

94. H. Duval, K. Rousseau, G. Eliès, et al., *Cloning, characterization, and comparative activity of turbot IGF-I and IGF-II*, *General and Comparative Endocrinology*, 2002, 126, 269-278.
95. Y. Yuan, Y. Hong, *Medaka insulin-like growth factor-2 supports self-renewal of the embryonic stem cell line and blastomeres in vitro*, *Scientific Reports*, 2017, 7, 78.
96. F.G. Ayson, E.G. de Jesus, S. Moriyama, et al., *Differential expression of insulin-like growth factor I and II mRNAs during embryogenesis and early larval development in rabbitfish, *Siganus guttatus**, *General and Comparative Endocrinology*, 2002, 126, 165-174.
97. J.B. Allard, C. Duan, *IGF-Binding Proteins: Why Do They Exist and Why Are There So Many?*, *Frontiers in Endocrinology*, 2018, 9, 117.
98. G.M. Neumann, L.A. Bach, *The N-terminal disulfide linkages of human insulin-like growth factor-binding protein-6 (hIGFBP-6) and hIGFBP-1 are different as determined by mass spectrometry*, *Journal of Biological Chemistry*, 1999, 274, 14587-14594.
99. L. Sun, S. Liu, R. Wang, et al., *Identification and Analysis of Genome-Wide SNPs Provide Insight into Signatures of Selection and Domestication in Channel Catfish (*Ictalurus punctatus*)*, *PLoS One*, 2014, 9, e109666.
100. H. Giral, U. Landmesser, A. Kratzer, *Into the Wild: GWAS Exploration of Non-coding RNAs*, *Frontiers in Cardiovascular Medicine*, 2018, 5, 181.
101. S. Hubert, B. Higgins, T. Borza, et al., *Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*)*, *BMC Genomics*, 2010, 11, 191.
102. E. Vallejos-Vidal, S. Reyes-Cerpa, J.A. Rivas-Pardo, et al., *Single-Nucleotide Polymorphisms (SNP) Mining and Their Effect on the Tridimensional Protein Structure Prediction in a Set of Immunity-Related Expressed Sequence Tags (EST) in Atlantic Salmon (*Salmo salar*)*, *Frontiers in genetics*, 2020, 10, 1406.
103. E. Diopere, B. Hellemans, F.A. Volckaert, et al., *Identification and validation of single nucleotide polymorphisms in growth- and maturation-related candidate genes in sole (*Solea solea* L.)*, *Marine Genomics*, 2013, 9, 33-38.

104. A.L. Ferchaud, M. Laporte, C. Perrier, et al., *Impact of supplementation on deleterious mutation distribution in an exploited salmonid*, *Evolutionary Applications*, 2018, 11, 1053-1065.
105. D. Suárez-Salgado, G.M. Parra-Bracamonte, F. Benavides-González, et al., *Non-synonymous polymorphisms in candidate gene associated with growth traits in Channel catfish (*Ictalurus punctatus*, Rafinesque, 1818)*, *Molecular Biology Reports*, 2020, 47, 87-95.
106. D. Robledo, J.A. Rubiolo, S. Cabaleiro, et al., *Differential gene expression and SNP association between fast- and slow-growing turbot (*Scophthalmus maximus*)*, *Scientific Reports*, 2017, 7, 12105.
107. M. Vera, J.-A. Alvarez-Dios, C. Fernandez, et al., *Development and validation of single nucleotide polymorphisms (SNPs) markers from two transcriptome 454-runs of turbot (*Scophthalmus maximus*) using high-throughput genotyping*, *International Journal of Molecular Sciences*, 2013, 14, 5694-5711.
108. S. Karlsson, T. Moen, S. Lien, et al., *Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip*, *Molecular Ecology Resources*, 2011, 11, 247-253.
109. C. Li, G. Waldbieser, B. Bosworth, et al., *SNP discovery in wild and domesticated populations of blue catfish, *Ictalurus furcatus*, using genotyping-by-sequencing and subsequent SNP validation*, *Molecular Ecology Resources*, 2014, 14, 1261-1270.
110. Y. Zhou, H.-C. Fu, Y.-Y. Wang, et al., *Genome-wide association study reveals growth-related SNPs and candidate genes in mandarin fish (*Siniperca chuatsi*)*, *Aquaculture*, 2022, 550, 737879.
111. P. Rotwein, *The insulin-like growth factor 2 gene and locus in nonmammalian vertebrates: Organizational simplicity with duplication but limited divergence in fish*, *Journal of Biological Chemistry*, 2018, 293, 15912-15932.
112. S. Varma Shrivastav, A. Bhardwaj, K.A. Pathak, et al., *Insulin-Like Growth Factor Binding Protein-3 (IGFBP-3): Unraveling the Role in Mediating IGF-Independent Effects Within the Cell*, *Frontiers in Cell and Developmental Biology*, 2020, 8, 286.

113. G.A. Luther, J. Lamplot, X. Chen, et al., *IGFBP5 domains exert distinct inhibitory effects on the tumorigenicity and metastasis of human osteosarcoma*, *Cancer Letters*, 2013, 336, 222-230.
114. *Insulin-like growth factor-binding protein 7 - UniProtKB - F8WH23 - Mouse*. <https://www.uniprot.org/uniprot/F8WH23>.
115. X.B. Fang, S.C. Liu, Q.Y. Wu, et al., *Linkage analysis of SNPs in IGFBP-6 and its relation with the body sizes of pig*, *Genetics and Molecular Research*, 2015b, 14, 17273-17280.
116. M.P. Mullen, D.P. Berry, D.J. Howard, et al., *Single Nucleotide Polymorphisms in the Insulin-Like Growth Factor 1 (IGF-1) Gene are Associated with Performance in Holstein-Friesian Dairy Cattle*, *Frontiers in Genetics*, 2011, 2, 3.
117. M.A. El-Magd, A.A. Saleh, A.A. Nafeaa, et al., *Polymorphisms of the IGF1 gene and their association with growth traits, serum concentration and expression rate of IGF1 and IGF1R in buffalo*, *Journal of Zhejiang University. Science. B*, 2017, 18, 1064-1074.
118. B.M. vonHoldt, R.Y. Kartzinel, C.D. Huber, et al., *Growth factor gene IGF1 is associated with bill size in the black-bellied seedcracker *Pyrenestes ostrinus**, *Nature Communications*, 2018, 9, 4855.
119. J.S. Renes, J. van Doorn, A.C.S. Hokken-Koelega, *Current insights into the role of the Growth Hormone-Insulin-like Growth Factor system in short children born small for gestational age*, *Hormone Research in Paediatrics*, 2019, 92, 15-27.
120. Q. Wu, H. Yu, X. Fang, et al., *The association of haplotypes in IGFBP-3 gene promoter region and tissue expressions in three pig breeds*, *Animal Cells and Systems*, 2016, 20, 384-393.
121. F. Mehrabi, M. Khalesi, A. Farhadi, *Polymorphism of insulin-like growth factor binding protein-3 (IGFBP-3) gene associated with growth traits in *Cyprinus carpio L**, *Iranian Journal of Ichthyology*, 2015, 2, 172-176.
122. J.T. Ou, S.Q. Tang, D.X. Sun, et al., *Polymorphisms of three neuroendocrine-correlated genes associated with growth and reproductive traits in the chicken*, *Poultry Science*, 2009, 88, 722-727.

123. W. Wang, K. Ouyang, X. Shangguan, et al., *Association of porcine IGF binding protein-5 gene with meat quality*, *Biochemical Genetics*, 2010, 48, 257-265.
124. S. Xie, D. Niu, K. Wei, et al., *Polymorphisms in the FOXO gene are associated with growth traits in the Sanmen breeding population of the razor clam *Sinonovacula constricta**, *Aquaculture and Fisheries*, 2018, 3, 177-183.
125. L. Cheng ,Y.H. Sun, *Polymorphisms in a myostatin gene and associations with growth in a hybrid of *Culter alburnus* and *Ancherythroculter nigrocauda**, *Genetics and Molecular Research*, 2015, 14, 5615-5620.
126. G. Lin, N.M. Thevasagayam, Z.Y. Wan, et al., *Transcriptome Analysis Identified Genes for Growth and Omega-3/6 Ratio in Saline Tilapia*, *Frontiers in Genetics*, 2019, 10, 244.
127. H. Huang, Y. Wei, Z. Meng, et al., *Polymorphisms of Leptin-b Gene Associated with Growth Traits in Orange-Spotted Grouper (*Epinephelus coioides*)*, *International journal of molecular sciences*, 2014, 15, 11996-12006.
128. P.E. Ulloa, G. Rincón, A. Islas-Trejo, et al., *RNA sequencing to study gene expression and SNP variations associated with growth in zebrafish fed a plant protein-based diet*, *Marine Biotechnology*, 2015, 17, 353-363.
129. J.L. McCauley, S.J. Kenealy, E.H. Margulies, et al., *SNPs in Multi-species Conserved Sequences (MCS) as useful markers in association studies: a practical approach*, *BMC genomics*, 2007, 8, 266.
130. M.A. Earp, E.L. Goode, *Linkage Disequilibrium*, Springer, 2017, Berlin, Heidelberg
131. X. Li, J. Bai, Y. Hu, et al., *Genotypes, haplotypes and diplotypes of IGF-II SNPs and their association with growth traits in largemouth bass (*Micropterus salmoides*)*, *Molecular Biology Reports*, 2012, 39, 4359-4365.
132. S. Zhang, X. Li, X. Chen, et al., *Significant associations between prolactin gene polymorphisms and growth traits in the channel catfish (*Ictalurus punctatus Rafinesque, 1818*) core breeding population*, *Meta Gene*, 2019, 19, 32-36.

133. C.U. Braz, J.F. Taylor, T. Bresolin, et al., *Sliding window haplotype approaches overcome single SNP analysis limitations in identifying genes for meat tenderness in Nelore cattle*, BMC Genetics, 2019, 20, 8.
134. G.J. McKinney, J.E. Seeb, L.W. Seeb, *Managing mixed-stock fisheries: genotyping multi-SNP haplotypes increases power for genetic stock identification*, Canadian Journal of Fisheries and Aquatic Sciences, 2017, 74, 429-434.
135. X.H. Li, J.J. Bai, X. Ye, et al., *Polymorphisms in the 5' flanking region of the insulin-like growth factor I gene are associated with growth traits in largemouth bass *Micropterus salmoides**, Fisheries Science, 2009, 75, 351-358.
136. S. Jaser, M. Dias, A. Lago, et al., *Single nucleotide polymorphisms in the growth hormone gene of *Oreochromis niloticus* and their association with growth performance*, Aquaculture Research, 2017, 48, 5835-5845.
137. M. Salem, R.L. Vallejo, T.D. Leeds, et al., *RNA-Seq identifies SNP markers for growth traits in rainbow trout*, PLoS One, 2012, 7, e36264.
138. L. Luo, R. Huang, A. Zhang, et al., *Selection of growth-related genes and dominant genotypes in transgenic Yellow River carp *Cyprinus carpio* L*, Functional & Integrative Genomics, 2018, 18, 425-437.

PHỤ LỤC

Mục lục phụ lục

Phụ lục 1: Danh sách 140 cá thể cá tra nuôi (70 cá thể tăng trưởng chậm và 70 cá thể tăng trưởng nhanh) được sử dụng trong bộ mẫu kiểm nghiệm.

Phụ lục 2: Kết quả xác định chuỗi peptit tín hiệu và vị trí domain chức năng trên protein IGFBP-1, -2, -3, -5, -6 của cá tra nuôi.

Phụ lục 3: Độ tương đồng về trình tự các đoạn gen trên contig được giải mã bằng NGS và các đoạn gen tương ứng được giải mã bằng phương pháp Sanger

Phụ lục 4: Kết quả tách chiết 20 mẫu DNA tổng số thuộc bộ mẫu khởi tạo.

Phụ lục 5: Điện di đồ sản phẩm PCR của các đoạn gen đích được khuếch đại từ DNA tổng số của 20 mẫu thuộc bộ mẫu khởi tạo.

Phụ lục 6: Điện di đồ một số mẫu DNA tổng số thuộc bộ 140 mẫu kiểm nghiệm.

Phụ lục 7: Kết quả đo nồng độ và độ tinh sạch của DNA tổng số tách chiết được từ 140 cá thể cá tra nuôi (70 cá thể tăng trưởng chậm và 70 cá thể tăng trưởng nhanh) được sử dụng trong bộ mẫu kiểm nghiệm.

Phụ lục 8: Điện di đồ sản phẩm PCR của các đoạn gen đích được khuếch đại từ DNA tổng số của 140 mẫu thuộc bộ mẫu kiểm nghiệm.

Phụ lục 9: Kết quả genotyping 10 SNP được kiểm nghiệm trên 160 cá thể sử dụng trong luận án.

Phụ lục 10: So sánh trình tự protein IGF1 của cá tra nuôi và các loài cá khác.

Phụ lục 11: So sánh trình tự protein IGF2 của cá tra nuôi và các loài cá khác.

Phụ lục 12: So sánh trình tự protein IGF1R của cá tra nuôi và các loài cá khác.

Phụ lục 1: Danh sách 140 cá thể cá tra nuôi (70 cá thể tăng trưởng chậm và 70 cá thể tăng trưởng nhanh) được sử dụng trong bộ mẫu kiểm nghiệm.

STT	ID	Khối lượng (gr)	EBV	Loại	Gia đình	Ký hiệu mẫu
1	0416D52B61	707,2	-341,2	TTC	85	016
2	0418CBCFB9	309,8	-332,5	TTC	47	207
3	04179E10A1	585	-295,4	TTC	45	190
4	0418CC05EC	423,6	-291,6	TTC	201	067
5	0418CBCA37	528,2	-291,3	TTC	57	046
6	0418CBD433	431,2	-287,5	TTC	201	180
7	0418CBEAF0	858,2	-286,9	TTC	48	220
8	0416D555B3	435,6	-281,2	TTC	45	111
9	0418F810F1	346,4	-269,9	TTC	201	072
10	0418CBDF76	498,6	-267,9	TTC	201	009
11	0418CBF2EC	473,2	-257,3	TTC	201	053
12	0418F7D24D	677	-251,3	TTC	179	84
13	0418F8183F	420,2	-238,6	TTC	250	080
14	0418CC0D47	694,4	-221,6	TTC	196	137
15	170001960895	4412	-219,2	TTC	116	141
16	0418F8451E	691,4	-217,5	TTC	198	149
17	170001856229	6158	-215,9	TTC	116	119
18	0418F81CCC	727,6	-215,8	TTC	199	049
19	0418F7D24D	677	-251,3	TTC	179	135
20	0418CBF86A	690,6	-211	TTC	196	079
21	0418F7E42B	704	-208,9	TTC	199	192
22	0418F805B4	495,6	-200	TTC	222	096
23	0418F7FC1D	4440	-194,7	TTC	229	409
24	170002049774	6040	-191,8	TTC	116	418
25	0418C831F0	351,8	-190,2	TTC	301	182
26	0418F8000F	700,4	-180,6	TTC	199	007
27	0418CBD57C	698,6	-183,5	TTC	51	014
28	0418CBD031	447,4	-176,3	TTC	296	045

STT	ID	Khối lượng (gr)	EBV	Loại	Gia đình	Ký hiệu mẫu
29	0418F7E379	695,2	-175,3	TTC	200	176
30	0418CC5E52	715,2	-135,1	TTC	31	061
31	170002052222	7124	-170,1	TTC	162	017
32	0418C843D6	6152	-164,1	TTC	111	124
33	170001770021	6130	-163,2	TTC	61	185
34	170001822893	4890	-163	TTC	55	052
35	0418CC09F5	795,6	-162	TTC	195	155
36	0418CBF5A0	968	-156,4	TTC	57	133
37	0418CBFE25	834,1	-152	TTC	200	191
38	0418CBBF8B	941,6	-151,4	TTC	200	150
39	0418CBF0B7	705,2	-148,1	TTC	222	082
40	0416D54FA4	530,6	-146,6	TTC	31	055
41	0418CC0B25	870	-145,5	TTC	196	062
42	0418F7DAB6	749,2	-143,4	TTC	200	188
43	0418F7D64F	741	-141,9	TTC	198	077
44	0418CBF6F6	1158,6	-141,6	TTC	196	040
45	0418F7D799	854,8	-140,5	TTC	200	189
46	0418CBC36E	659,2	-138,8	TTC	174	162
47	0418F7C992	745,2	-138	TTC	199	112
48	04179136B7	526,6	-136,7	TTC	31	047
49	0418CC00BE	574,6	-136,1	TTC	249	144
50	0418C84FB4	430,2	-136	TTC	174	136
51	0418F7CE3A	768	-135,7	TTC	199	076
52	0418CC5E52	715,2	-135,1	TTC	31	125
53	0418F7E2C2	774,4	-130,6	TTC	200	146
54	0418CBE050	993,6	-129,4	TTC	200	120
55	0418CBF2CF	831,2	-126,3	TTC	200	209
56	0418CBF7DE	590	-125,9	TTC	296	211
57	0418F80237	787,8	-124,7	TTC	222	199
58	0418CBDEDB	773,8	-123,4	TTC	222	186
59	0418CBCC10	729	-121,8	TTC	249	202

STT	ID	Khối lượng (gr)	EBV	Loại	Gia đình	Ký hiệu mẫu
60	0418CBC992	607,6	-121,2	TTC	222	212
61	0418F7CC77	665,2	-121	TTC	222	131
62	0418CBE555	907,8	-120,2	TTC	200	179
63	0418CBD6F2	800,8	-119,6	TTC	201	148
64	0418CC0FCE	432,6	-119,4	TTC	296	173
65	0418CC0E56	473,4	-119,2	TTC	174	215
66	0418F81D38	796	-117,9	TTC	198	075
67	0418CBC9A9	906,4	-117,8	TTC	201	138
68	0418CBC3D1	589,4	-114,2	TTC	242	054
69	0418CC0717	845,6	-113,5	TTC	222	034
70	0416D57970	677,8	-100,6	TTC	296	069
71	0418CC0B29	981,5	84,7	TTN	191	101
72	0418CBFD5B	1248,6	213,2	TTN	149	097
73	0418CBE0BD	1307,6	221,8	TTN	149	134
74	0418CBE8B1	1603,2	242	TTN	191	159
75	0418CBDE1B	1768,6	278,1	TTN	154	216
76	0418CBBE75	1815,4	281,9	TTN	154	092
77	0418CC1122	1624	286,7	TTN	154	194
78	0418CC45B1	1354,6	287,6	TTN	269	063
79	0418CC0484	1293,2	287,6	TTN	105	167
80	0418CBBF3B	1345,8	295,7	TTN	105	078
81	0418CBF5C1	1364,4	300,5	TTN	105	107
82	0418F7FC7F	1247,4	302,1	TTN	271	023
83	04179A1D97	1367,4	303,8	TTN	105	030
84	0417BD3998	25005	305	TTN	11	020
85	0418C7C760	1307,6	305,6	TTN	271	201
86	0417D8752A	21866	308,9	TTN	30	411
87	0418C7D957	1303,2	309	TTN	271	057
88	0418CBDC9A	1873,6	310,6	TTN	154	197
89	0418CC23BC	1413,4	311,3	TTN	38	195
90	0418CC0A03	1506	314,1	TTN	110	166

STT	ID	Khối lượng (gr)	EBV	Loại	Gia đình	Ký hiệu mẫu
91	0418CBE46D	1365,6	316,1	TTN	180	122
92	170002054001	18290	320,8	TTN	10	170
93	170001820441	21438	323,5	TTN	269	412
94	0418C7C0F3	1326,6	323,6	TTN	271	011
95	0418CBF29E	1486,6	323,6	TTN	105	088
96	0418CC04C5	1432,4	324,8	TTN	180	177
97	0418CBF227	1950,2	326,3	TTN	10	121
98	0418CBFB3F	1341,2	327,3	TTN	110	066
99	0418CBFA0D	1585,6	328,8	TTN	149	175
100	0418CBD78B	1608,4	329,4	TTN	173	068
101	0418CBC351	1568,6	329,4	TTN	180	074
102	0418CBDEB0	1439	331,6	TTN	152	157
103	0418CC083A	1494,2	334	TTN	131	018
104	0418CBED42	1422,2	334,9	TTN	180	012
105	0418CBFAAE	1394,6	335,6	TTN	269	123
106	0418CC4729	2188	337,5	TTN	10	038
107	0418CC0CF7	1352,4	338,6	TTN	105	158
108	0418CBE696	1568,2	343,3	TTN	173	218
109	0418C7CBC9	1367,2	343,8	TTN	271	083
110	0418CBCE50	1353,2	345,2	TTN	101	035
111	0418CBEB95	1962,2	345,7	TTN	175	204
112	0418F7F836	1386	346,8	TTN	269	118
113	0418CBEFDB	1312,8	348,6	TTN	101	168
114	0418CBD6EF	1599,2	351,5	TTN	131	042
115	0418CBC8E9	1992,4	351,8	TTN	175	183
116	0418CBC037	1566	354,3	TTN	46	181
117	0418CC132A	1457,6	357,6	TTN	173	031
118	0417BFAF20	25384	357,9	TTN	38	094
119	0418C81B18	1434,2	360,4	TTN	241	172
120	0418CBC88A	1924,6	360,7	TTN	95	021
121	0418CBEAEA	1677,6	363,7	TTN	131	130

STT	ID	Khối lượng (gr)	EBV	Loại	Gia đình	Ký hiệu mẫu
122	0416D539B6	1375,3	363,9	TTN	117	205
123	0418CC138E	1785,4	366,2	TTN	161	102
124	0418CBEFDB	1312,8	348,6	TTN	101	043
125	0418CBDCAB	2267,6	367,8	TTN	154	013
126	0418CBFCC8	1388	379,5	TTN	30	200
127	0418CBF36B	1353	382,1	TTN	101	003
128	0418CC1056	1694	382,5	TTN	46	024
129	0418CBFD3B	1467,6	383,1	TTN	101	002
130	0418CBDE1B	1768,6	278,1	TTN	154	090
131	0418CBFD5B	1248,6	213,2	TTN	149	041
132	0418CC0003	1709,6	416,7	TTN	173	151
133	0418CBEC42	1802,2	418,6	TTN	101	059
134	0418CBC3F8	1970,6	419,1	TTN	113	143
135	0418CC0CE8	1802,2	428,5	TTN	180	165
136	0418CBD6DA	2598	449,7	TTN	175	109
137	0418CBC4D4	1713,4	458	TTN	153	142
138	0418CBD171	2205,4	464,2	TTN	113	213
139	0418CC029B	2157,6	485	TTN	113	145
140	0418CBF51E	2056	514,2	TTN	153	128

Phụ lục 2: Kết quả xác định chuỗi peptit tín hiệu và vị trí domain chức năng trên protein IGFBP-1, -2, -3, -5, -6 của cá tra nuôi.

Protein	Vị trí cắt của enzyme trên chuỗi peptide tín hiệu trên protein tương ứng	Vị trí domain chức năng trên protein tương ứng (Từ axit amin đến axit amin)	
		Domain IGFBP	Domain thyroglobulin type 1
IGFBP-1	Giữa axit amin 25 và 26 (LRA-SP)	32-109	184-259
IGFBP-2	Giữa axit amin 20 và 21 (ARA-EL)	24-102	178-258
IGFBP-3	Giữa axit amin 34 và 35 (VHA-AG)	39-116	224-285
IGFBP-5	Giữa axit amin 21 và 22 (CSG-SY)	24-100	188-259
IGFBP-6	Giữa axit amin 31 và 32 (TLA-SR)	40-101	130-201

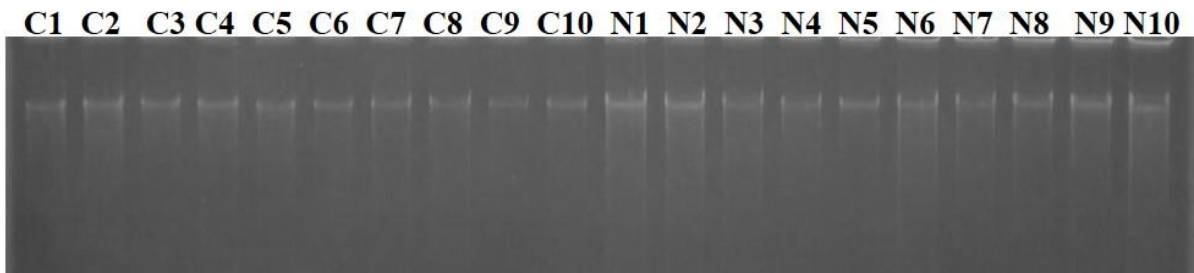
Phụ lục 3: Độ tương đồng về trình tự các đoạn gen trên contig được giải mã bằng NGS và các đoạn gen tương ứng được giải mã bằng phương pháp Sanger

Tên contig^(*)/ vị trí đoạn trên contig	Tên gen/ Vị trí đoạn trên gen	Chiều dài đoạn / Độ tương đồng
sc0000011 NW_020824206.1(8210909..8211445)	<i>IGF1</i> / -130 đến 407	537 bp/ 100 %
sc0000011 NW_020824206.1 (8224108..8225013)	<i>IGF1</i> / 13070 đến 13975	906 bp/ 100 %
sc0000011 NW_020824206.1 (8227236..8228466)	<i>IGF1</i> / 16198 đến 17428	1231 bp/ 99,9 %
sc0000028 NW_020824223 (1587043..1590549)	<i>IGF2</i> / 49 đến 3555	3507 bp/ 99,9 %
sc0000015 NW_020824210.1 (3329301..3333717)	<i>IGF1R</i> / 12625 đến 17041	4417 bp/ 99,9 %
sc0000015 NW_020824210.1 (3400218..3403298)	<i>IGF1R</i> / 83542 đến 86622	3081 bp/ 99,9 %
sc0000015 NW_020824210.1 (3403369..3408217)	<i>IGF1R</i> / 86693 đến 91541	4849 bp/ 100 %
sc0000042 NW_020824237.1 (1915637..1918352)	<i>IGFBP-1</i> / -308 đến 2408	2716 bp/ 99,9 %
sc0000046 NW_020824241.1 (4426946..4428118)	<i>IGFBP-2</i> / -129 đến 1044	1173 bp/ 99,9 %
sc0000046 NW_020824241.1(4410046..4411118)	<i>IGFBP-2</i> / 16872 đến 17943	1072 bp/ 100 %
sc0000046 NW_020824241.1 (4399846..4404377)	<i>IGFBP-2</i> / 23613 đến 28144	4532 bp/ 99,9 %
sc0000042 NW_020824237.1 (1938399..1939421)	<i>IGFBP-3</i> / 267 đến 1287	1021 bp/ 99,9 %
sc0000042 NW_020824237.1 (1932959..1935177)	<i>IGFBP-3</i> / 4511 đến 6729	2219 bp/ 100 %
sc0000042 NW_020824237.1 (1922386..1923436)	<i>IGFBP-3</i> / 16252 đến 17302	1051 bp/ 100 %
sc0000003 NW_020824198.1 (9142508..9144067)	<i>IGFBP-5</i> / -123 đến 1437	1560 bp/ 100 %
sc0000003 NW_020824198.1 (9149736..9152058)	<i>IGFBP-5</i> / 7106 đến 9428	2323 bp/ 100 %

Tên contig ^(*)/ vị trí đoạn trên contig	Tên gen/ Vị trí đoạn trên gen	Chiều dài đoạn / Độ tương đồng
sc0000003 NW_020824198.1 (9153367..9154872)	<i>IGFBP-5</i> / 10737 đến 12242	1506 bp/ 100 %
sc0000009 NW_020824204.1 (3856187..3860501)	<i>IGFBP-6</i> / -493 đến 3822	4315 bp/ 100 %
sc0000054 NW_020824249.1 (1274146..1277518)	<i>IGFBP-7</i> / -516 đến 2857	3373 bp/ 99,9 %
sc0000054 NW_020824249.1 (1270374..1273405)	<i>IGFBP-7</i> / 3598 đến 6629	3032 bp/ 100 %
sc0000054 NW_020824249.1 (1267956..1269690)	<i>IGFBP-7</i> / 7313 đến 9047	1735 bp/ 99,9 %

(*): Theo trình tự hệ gen cá tra nuôi giải mã bằng NGS [16].

Phụ lục 4: Kết quả tách chiết 20 mẫu DNA tổng số thuộc bộ mẫu khởi tạo.



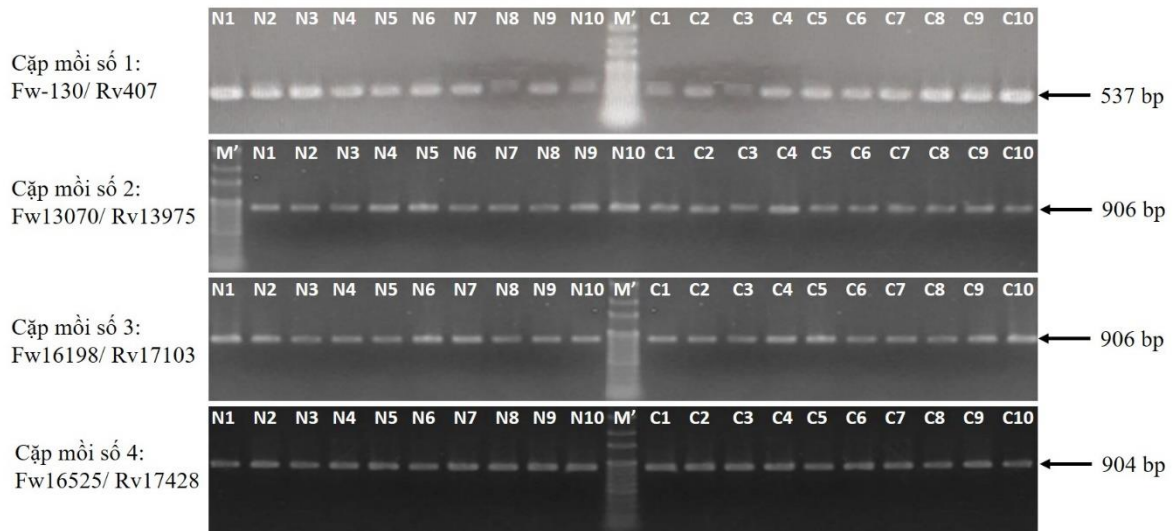
Điện di đồ 20 mẫu DNA tổng số trên gel agarose 1%.

C1 đến C10: DNA tổng số của 10 mẫu cá tăng trưởng chậm. N1 đến N10: DNA tổng số của 10 mẫu cá tăng trưởng nhanh.

Kết quả đo nồng độ và độ tinh sạch của DNA tổng số tách chiết được từ 20 cá thể trên (10 cá thể tăng trưởng nhanh (TTN) và 10 cá thể tăng trưởng chậm (TTC))

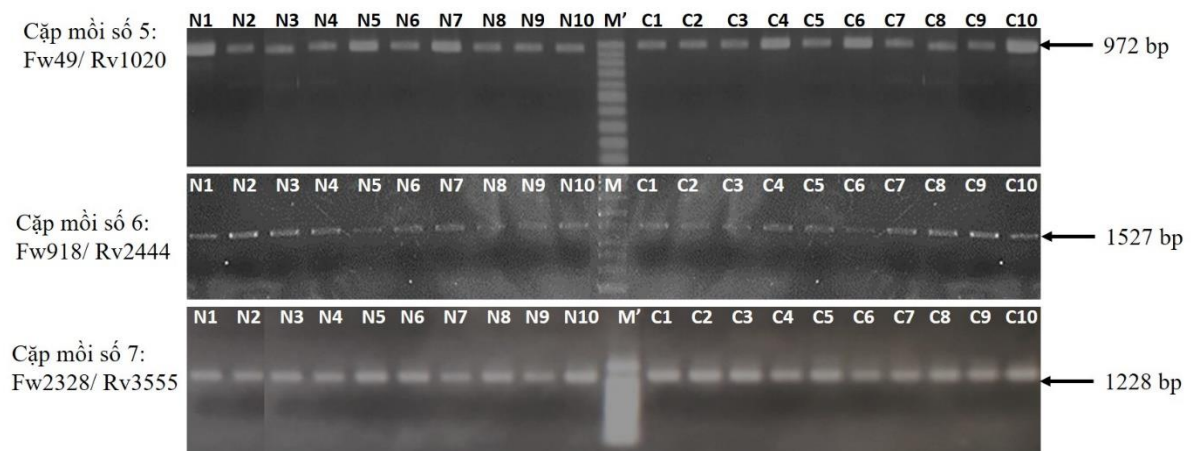
STT	Mẫu TTN	Nồng độ (ng/μl)	OD (A260/A280)	STT	Mẫu TTC	Nồng độ (ng/μl)	OD (A260/A280)
1	N1	1214	1,85	1	C1	402	2,06
2	N2	222	2,03	2	C2	435	2,09
3	N3	374	2,04	3	C3	288	2,08
4	N4	435	2,08	4	C4	340	2,07
5	N5	252	2,07	5	C5	196	2,08
6	N6	399	1,78	6	C6	239	1,92
7	N7	352	1,99	7	C7	284	2,04
8	N8	1155	2,11	8	C8	415	2,09
9	N9	536	2,09	9	C9	196	2,05
10	N10	350	1,83	10	C10	357	1,93

Phụ lục 5: Điện di đồ sản phẩm PCR của các đoạn gen đích được khuếch đại từ DNA tổng số của 20 mẫu thuộc bộ mẫu khởi tạo



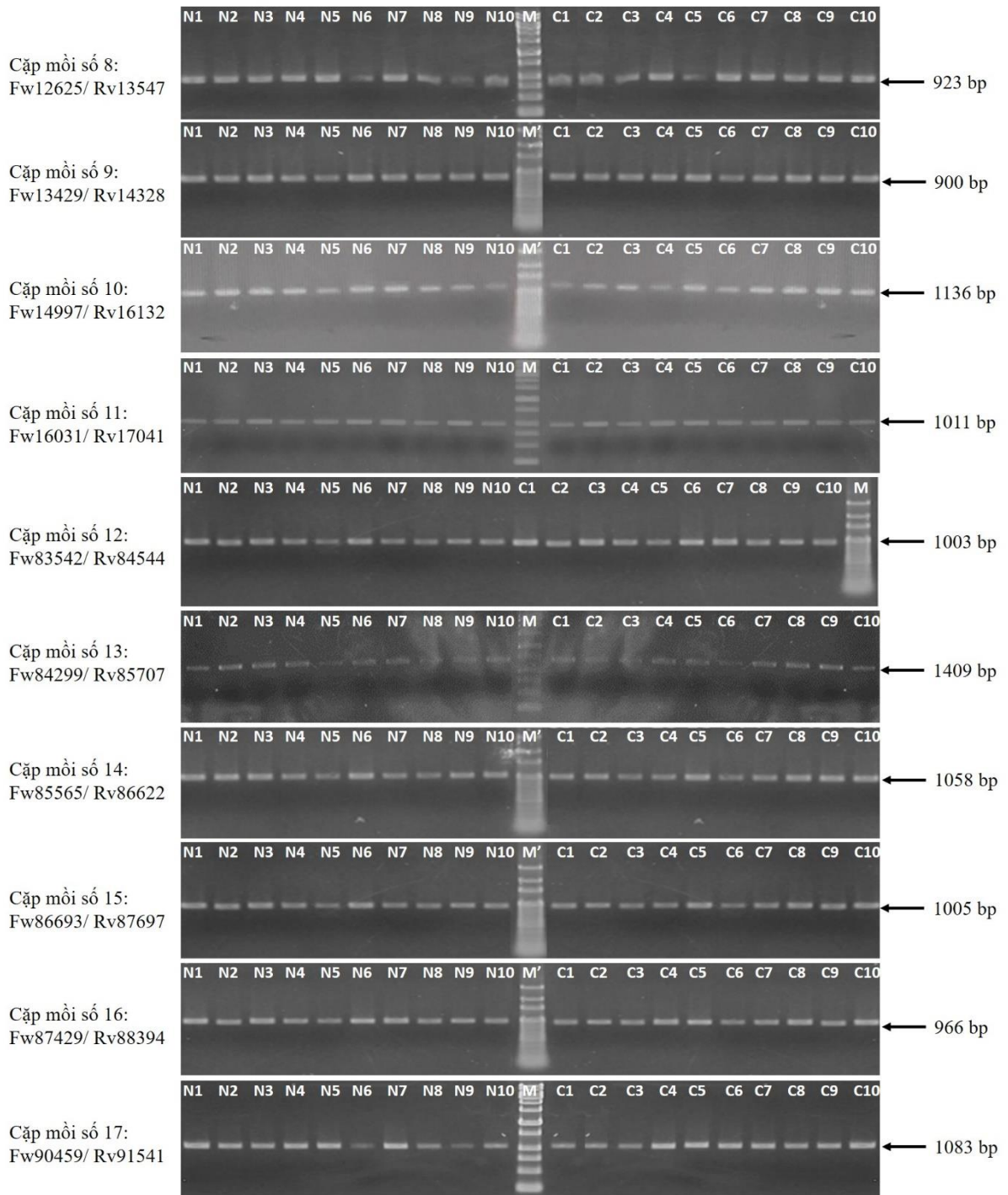
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGF1*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



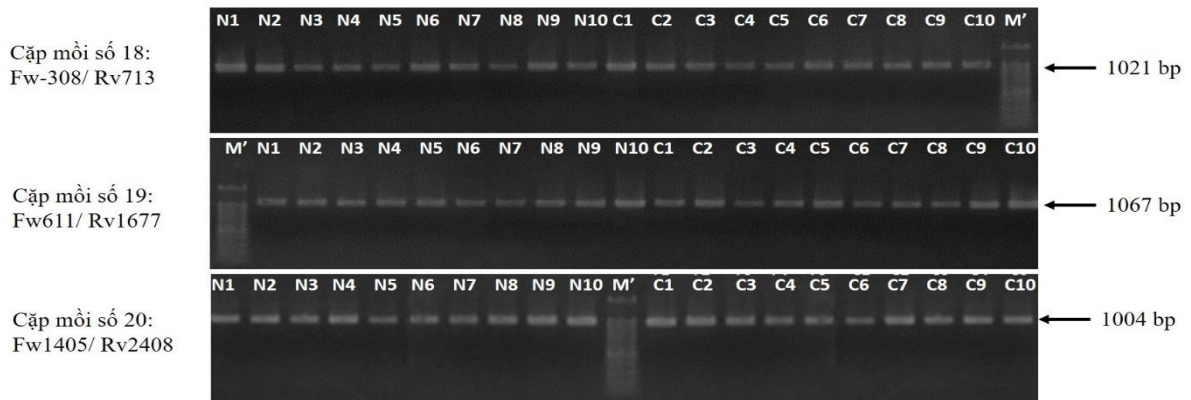
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGF2*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGF1R*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



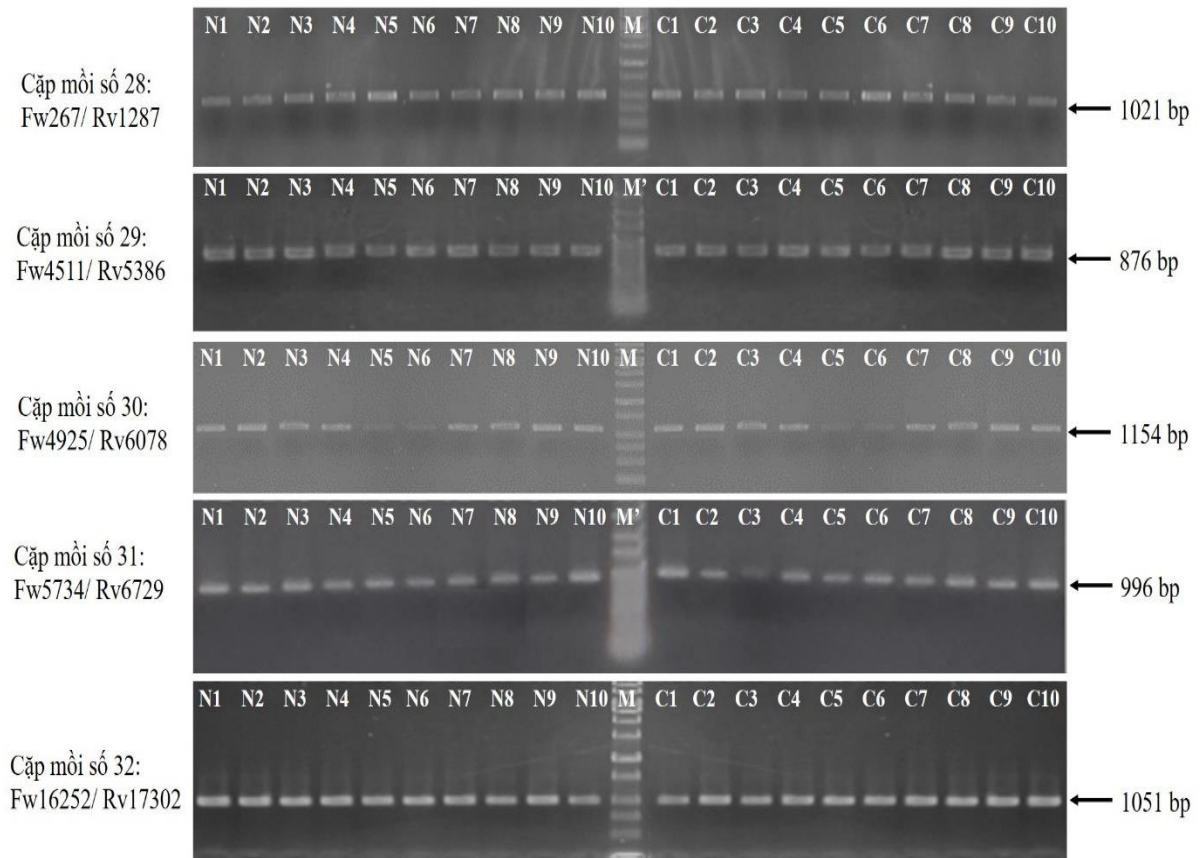
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-1*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



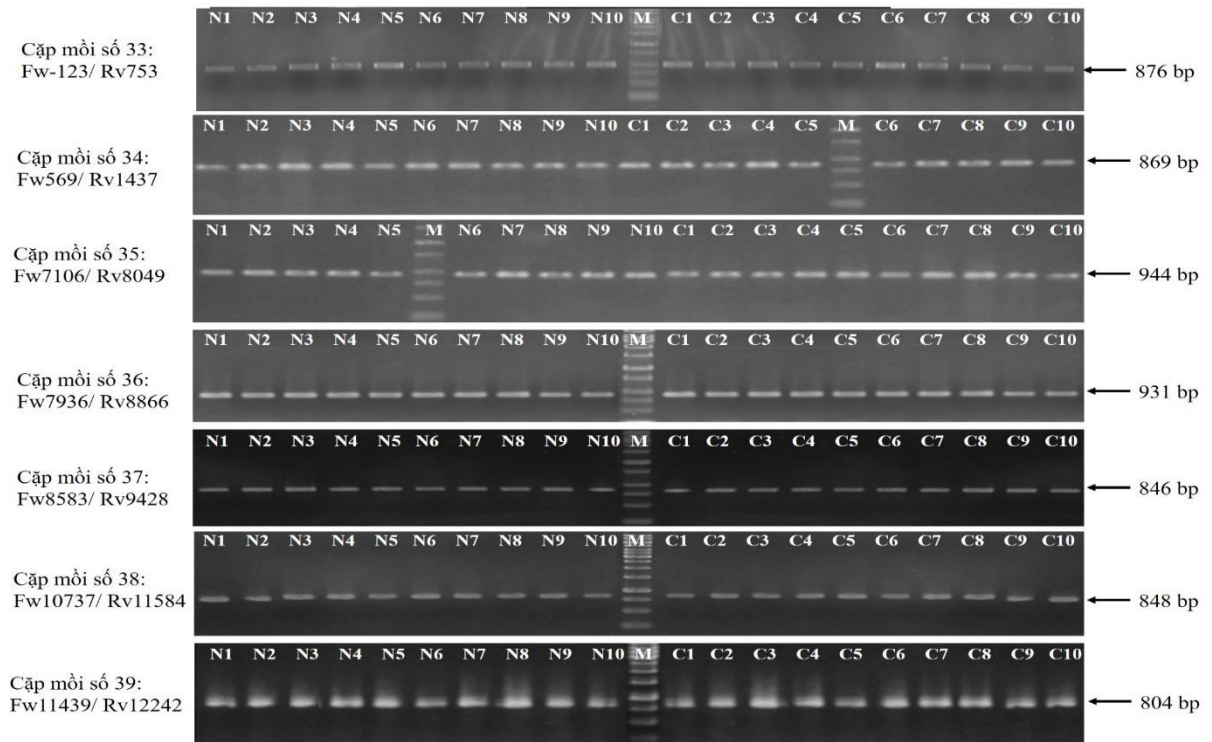
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-2*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



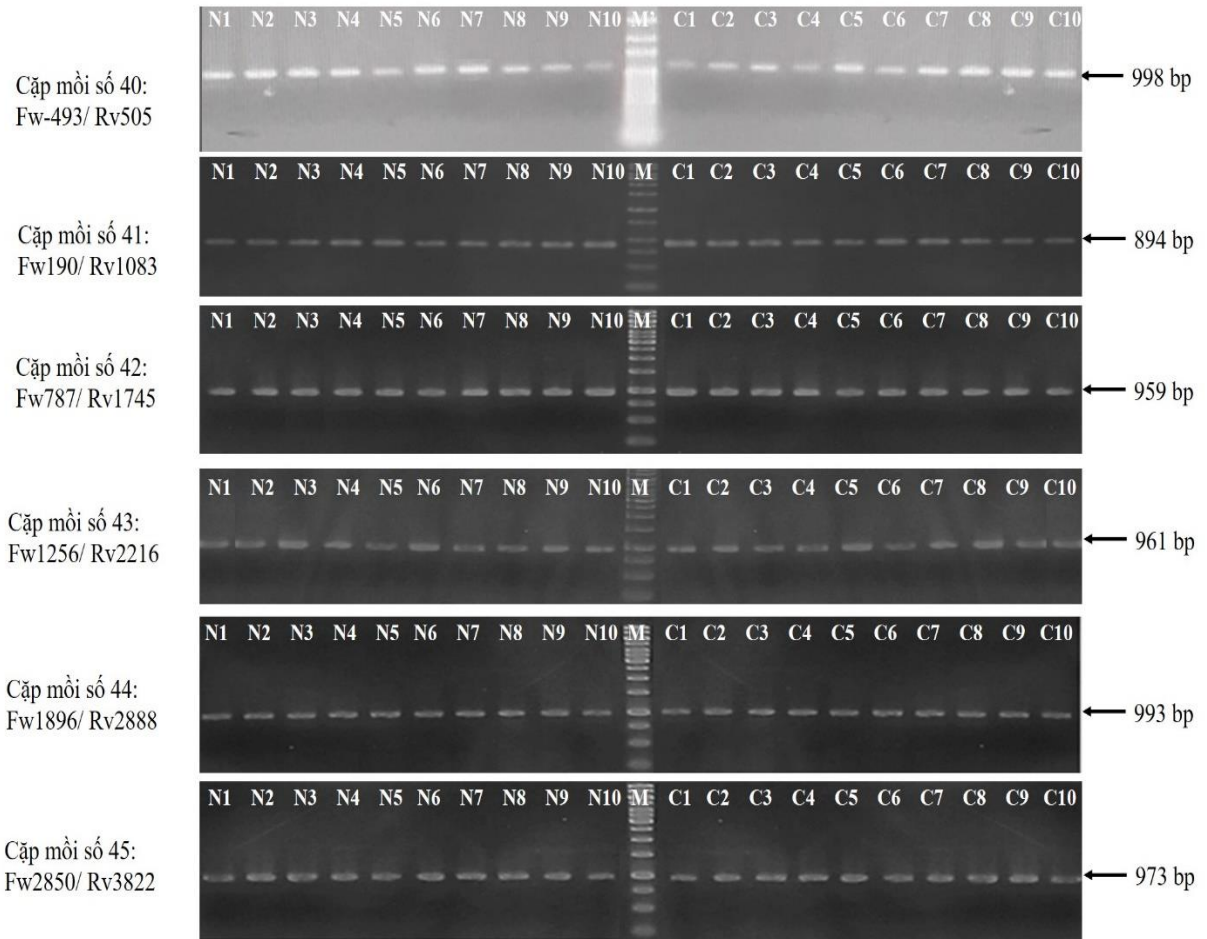
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-3*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



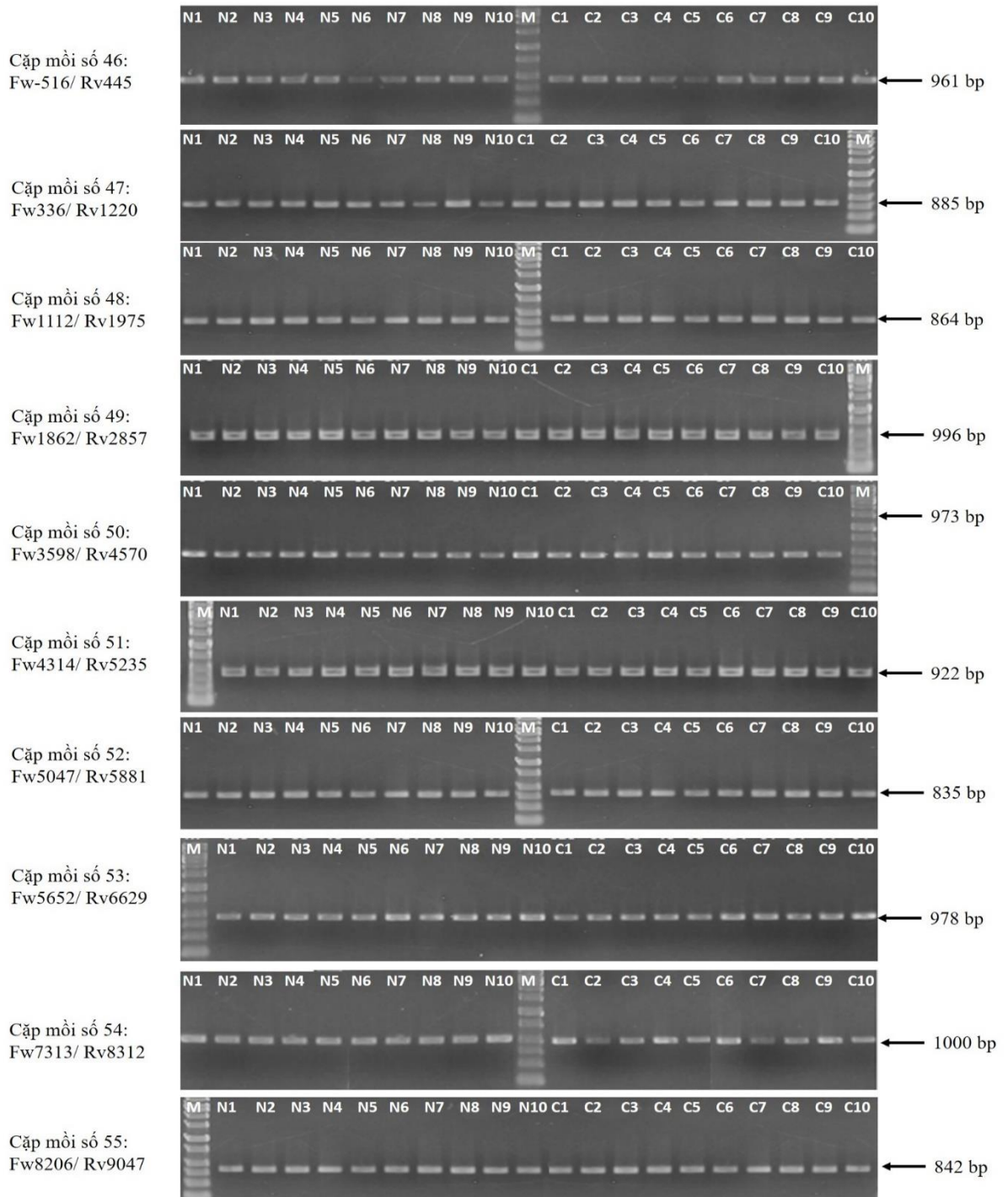
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-5*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



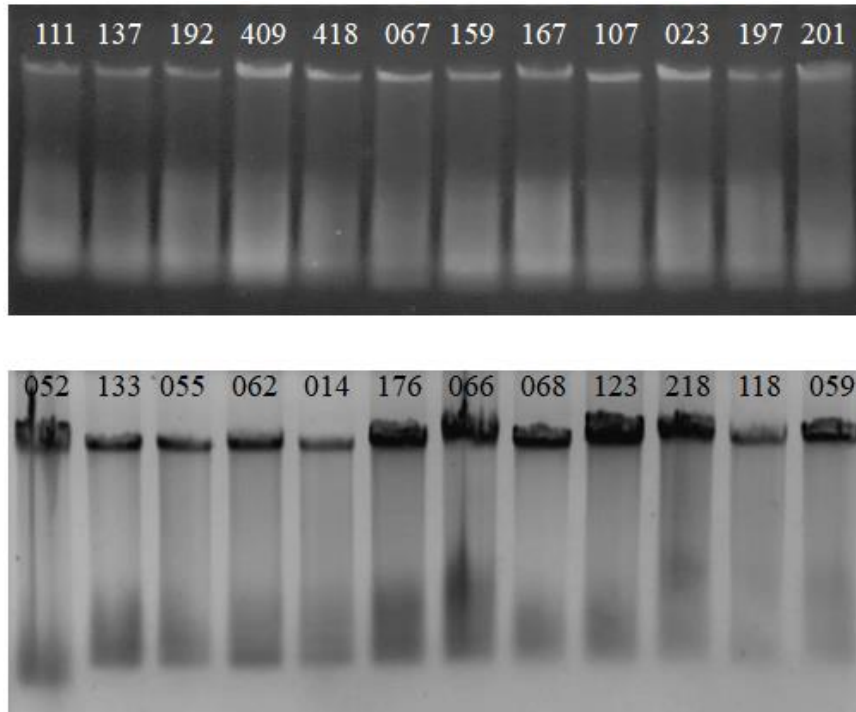
Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-6*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). M': Marker 100 bp (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.



Điện di đồ sản phẩm PCR khuếch đại các đoạn gen *IGFBP-7*

Từ N1 đến N10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng nhanh. Từ C1 đến C10 là sản phẩm PCR được nhân lên từ DNA tổng số của 10 cá thể cá tra tăng trưởng chậm. M: Marker 1 kb (Fermentas). Các mũi tên ở bên phải hình là vị trí các băng sản phẩm.

Phụ lục 6: Điện di đồ một số mẫu DNA tổng số thuộc bộ mẫu kiểm nghiệm

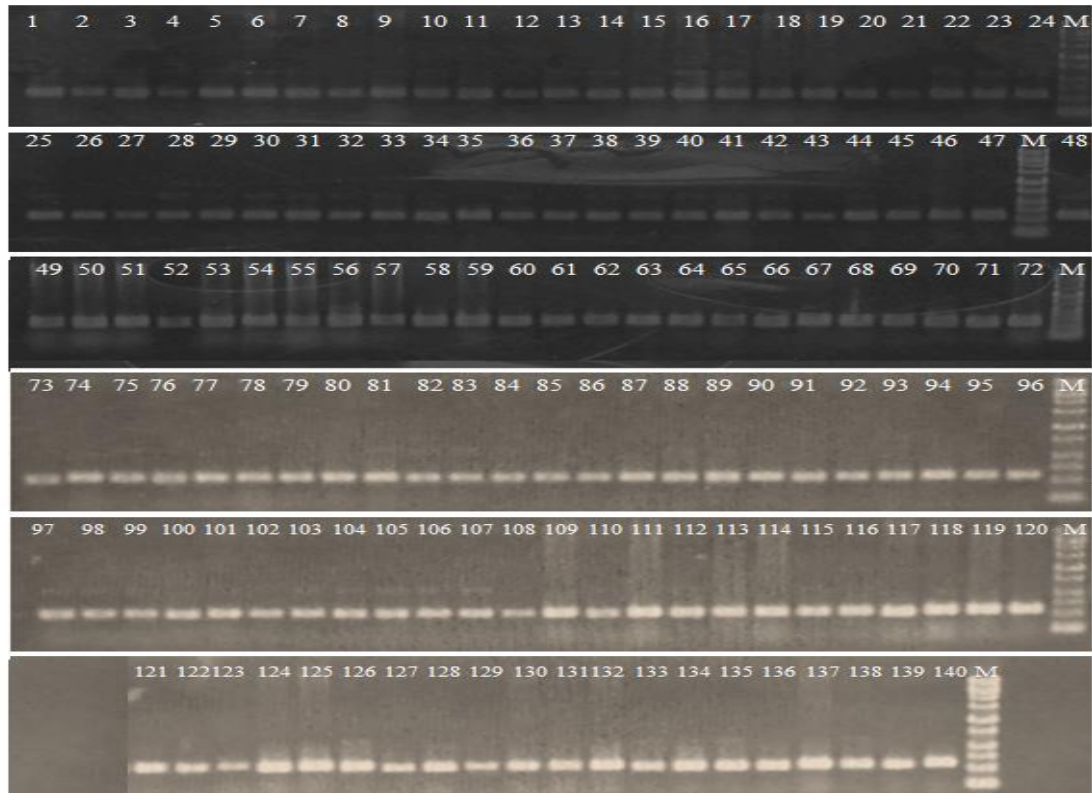
Điện di đồ một số mẫu DNA tổng số trên gel agarose 1%.
(Số trên mỗi giếng là số ký hiệu mẫu)

Phụ lục 7: Kết quả đo nồng độ và độ tinh sạch của DNA tổng số tách chiết được từ 140 cá thể cá tra nuôi (70 cá thể tăng trưởng chậm và 70 cá thể tăng trưởng nhanh) được sử dụng trong bộ mẫu kiểm nghiệm.

STT	Mẫu TTN	Nồng độ (ng/μl)	OD (A260/A280)	STT	Mẫu TTC	Nồng độ (ng/μl)	OD (A260/A280)
1	2	1516,9	1,84	1	7	1693,4	1,82
2	3	2554,9	1,75	2	9	1988,3	1,77
3	11	1002,4	1,83	3	14	1011,1	1,83
4	12	1295,3	1,84	4	16	1914,5	1,81
5	13	1658,3	1,85	5	17	1910,1	1,75
6	18	1696,3	1,82	6	34	2704,2	1,75
7	20	1463,2	1,82	7	40	1801,9	1,82
8	21	1932,5	1,80	8	45	1486,3	1,82
9	23	1876,5	1,84	9	46	1415,3	1,81
10	24	1316,4	1,87	10	47	1993,2	1,83
11	30	1997,5	1,76	11	49	2540,9	1,75
12	31	1354,3	1,89	12	52	2658,4	1,76
13	35	1356,8	1,88	13	53	1958,4	1,75
14	38	1856,4	1,85	14	54	2346,2	1,77
15	41	1352,9	1,86	15	55	2658,4	1,81
16	42	1728,6	1,86	16	61	2658,4	1,82
17	43	1059,8	1,84	17	62	2540,9	1,75
18	57	1540,9	1,85	18	67	2056,8	1,86
19	59	1530,6	1,83	19	69	2035,6	1,83
20	63	1889,4	1,82	20	72	2259,4	1,81
21	66	1658,4	1,84	21	75	2589,4	1,83
22	68	1468,2	1,87	22	76	2560,6	1,81
23	74	1069,2	1,87	23	77	1756,1	1,79
24	78	1463,8	1,85	24	79	2095,3	1,75
25	83	2053,8	1,81	25	80	1011,2	1,81
26	88	1968,6	1,81	26	82	2040,1	1,78
27	90	1509,4	1,79	27	84	1903,1	1,86
28	92	1560,6	1,86	28	96	2256,1	1,82
29	94	1530,2	1,81	29	111	2720,9	1,75
30	97	1595,3	1,88	30	112	1298,7	1,82
31	101	2017,3	1,81	31	119	2400,9	1,76
32	102	1456,8	1,84	32	120	2912,1	1,75
33	107	2006,4	1,89	33	124	2965,6	1,75
34	109	2090,3	1,85	34	125	2450,1	1,80
35	118	1325,6	1,82	35	131	2684,5	1,79
36	121	1865,4	1,75	36	133	2400,9	1,79

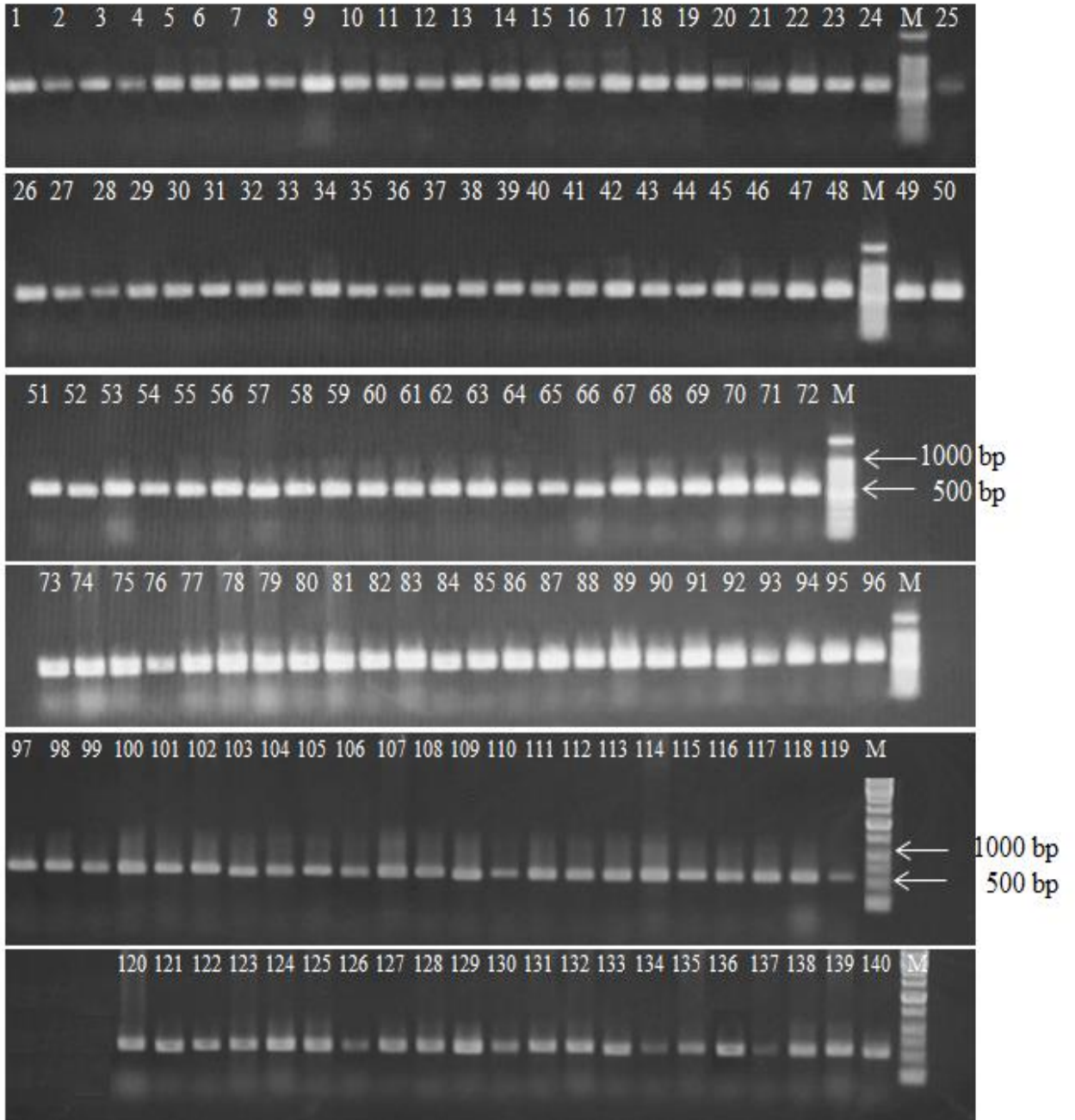
STT	Mẫu TTN	Nồng độ (ng/μl)	OD (A260/A280)	STT	Mẫu TTC	Nồng độ (ng/μl)	OD (A260/A280)
37	122	1958,3	1,87	37	135	2460,3	1,81
38	123	1980,6	1,89	38	136	2243,1	1,75
39	128	1563,5	1,83	39	137	1389,7	1,79
40	130	1496,8	1,84	40	138	2009,1	1,85
41	134	2661,3	1,77	41	141	2193,5	1,81
42	142	1060,3	1,81	42	144	1498,3	1,84
43	143	1264,1	1,83	43	146	2998,7	1,79
44	145	1763,7	1,86	44	148	2268,4	1,80
45	151	2035,2	1,81	45	149	2509,1	1,79
46	157	1615,2	1,76	46	150	2620,3	1,82
47	158	1669,1	1,82	47	155	2596,2	1,80
48	159	1445,3	1,83	48	162	1849,5	1,79
49	165	1356,3	1,85	49	173	2098,9	1,81
50	166	1298,3	1,76	50	176	399,5	1,85
51	167	1569,8	1,81	51	179	2540,9	1,81
52	168	1599,2	1,82	52	180	1740,1	1,79
53	170	1012,3	1,75	53	182	1790,9	1,82
54	172	1011,2	1,75	54	185	2689,1	1,79
55	175	1259,3	1,84	55	186	2560,8	1,80
56	177	2931,1	1,75	56	188	2478,9	1,79
57	181	2921,6	1,81	57	189	1750,1	1,82
58	183	1848,2	1,86	58	190	1648,2	1,82
59	194	1556,1	1,85	59	191	2635,3	1,81
60	195	1865,2	1,86	60	192	2056,2	1,75
61	197	1963,5	1,84	61	199	1865,2	1,81
62	200	1560,1	1,86	62	202	2648,2	1,79
63	201	1720,1	1,84	63	207	2556,1	1,79
64	204	1457,0	1,81	64	209	2865,2	1,76
65	205	1886,2	1,85	65	211	2457,0	1,75
66	213	1432,0	1,83	66	212	2486,2	1,82
67	216	1405,2	1,89	67	215	2256,1	1,86
68	218	1409,8	1,88	68	220	2269,2	1,76
69	411	1905,7	1,87	69	409	2965,2	1,78
70	412	2398,2	1,75	70	418	1550,3	1,75

Phụ lục 8: Điện di đồ sản phẩm PCR được khuếch đại từ DNA tổng số của 140 mẫu thuộc bộ mẫu kiểm nghiệm.

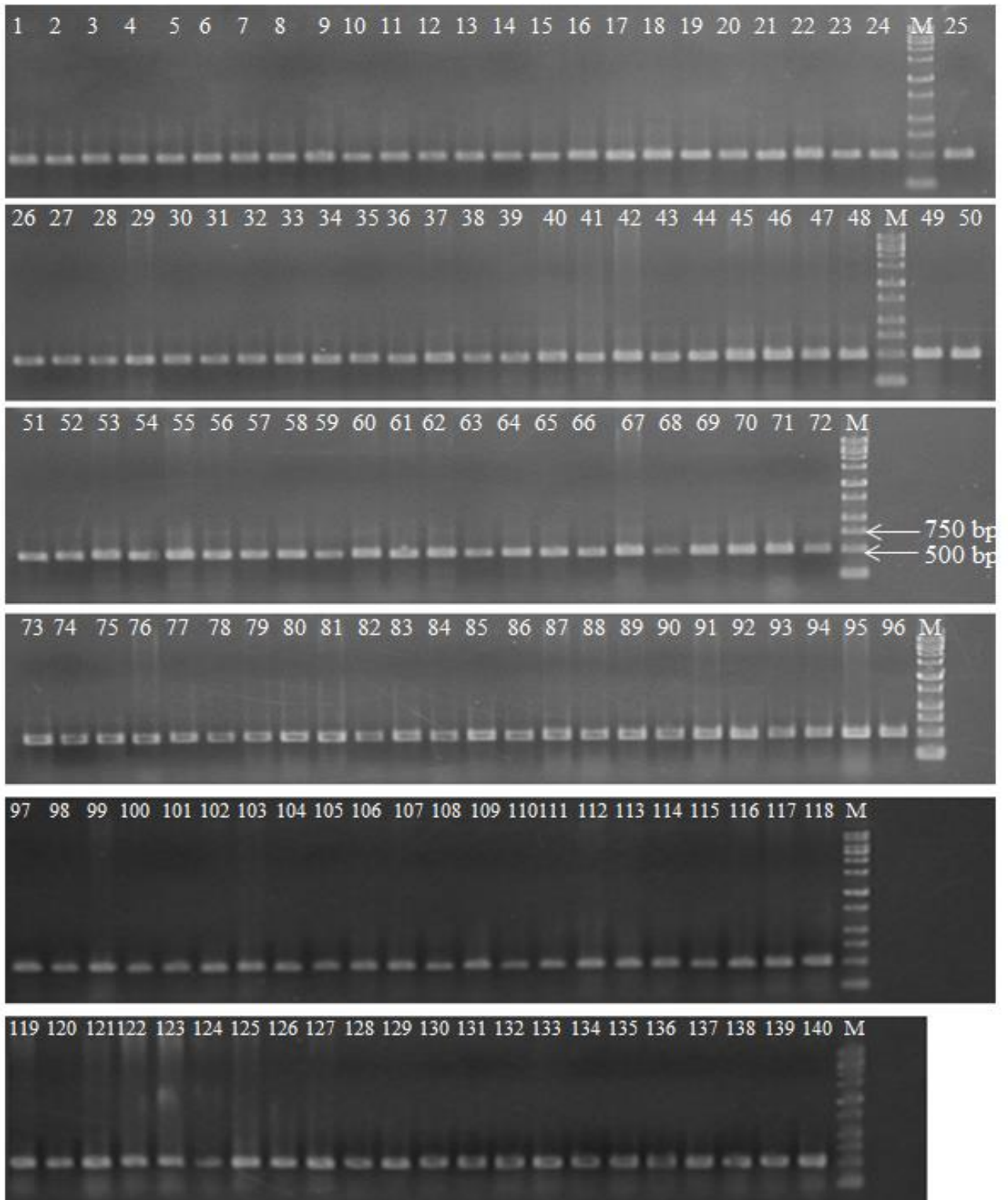


Kết quả điện di sản phẩm PCR khuếch đại từ cặp môi nhân đoạn dài 597 bp chứa SNP *IGF1*.13680 A>T từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.

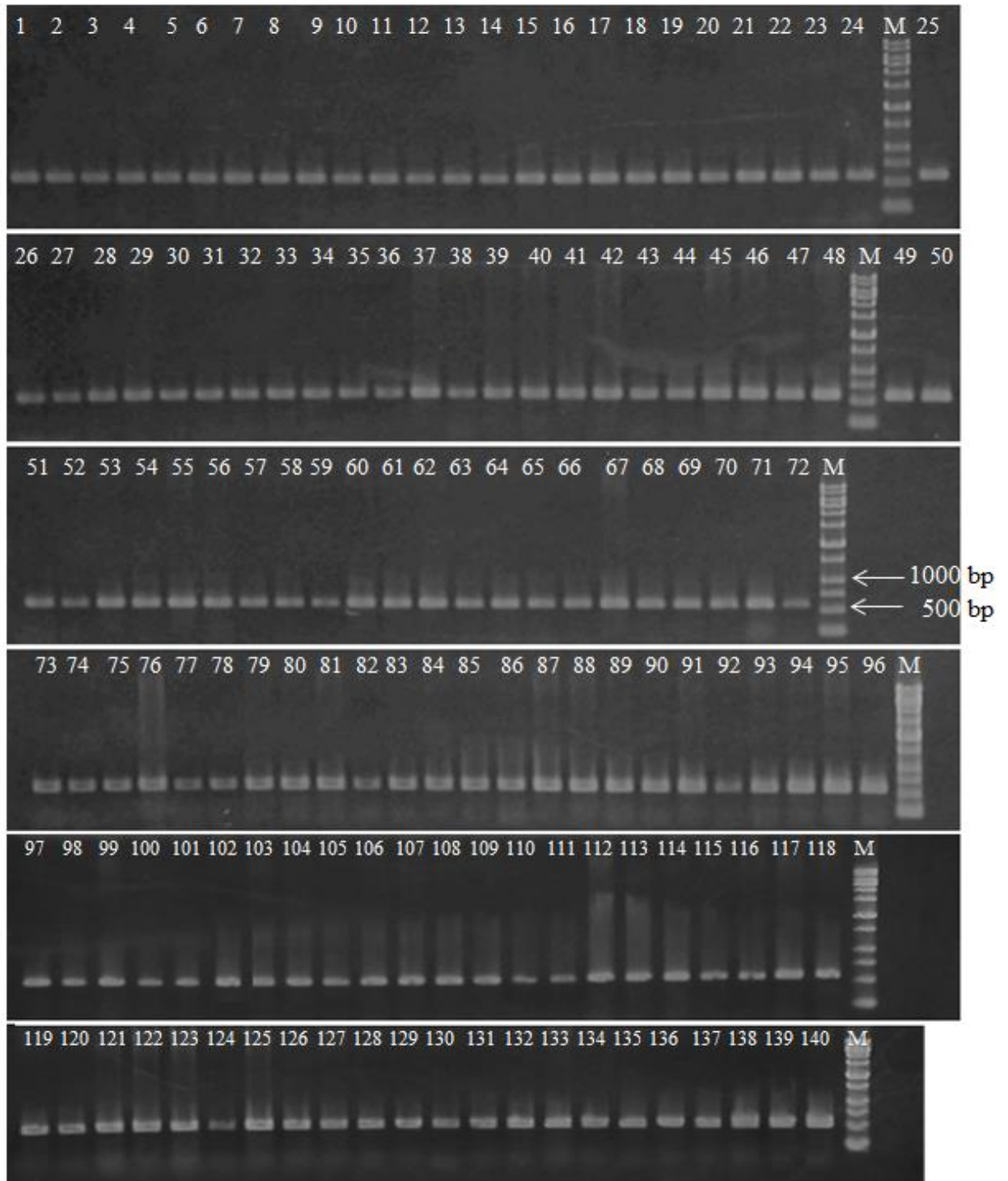
Số trên mỗi giếng là số thứ tự mẫu ở bảng Phụ lục 1. M: Marker 1 kb (Fermentas).



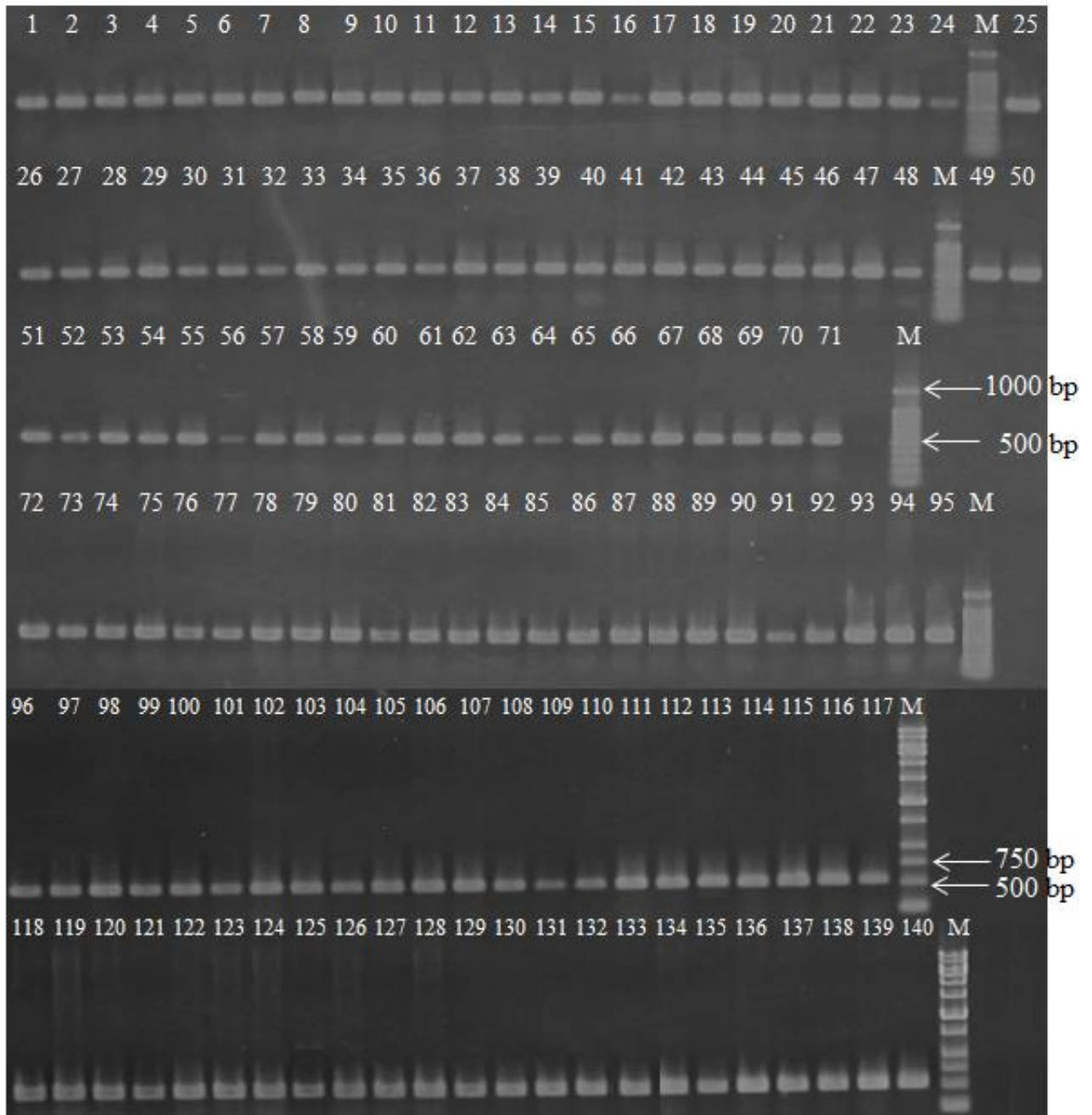
Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 595 bp chứa SNP *IGF1R.13357T>C* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



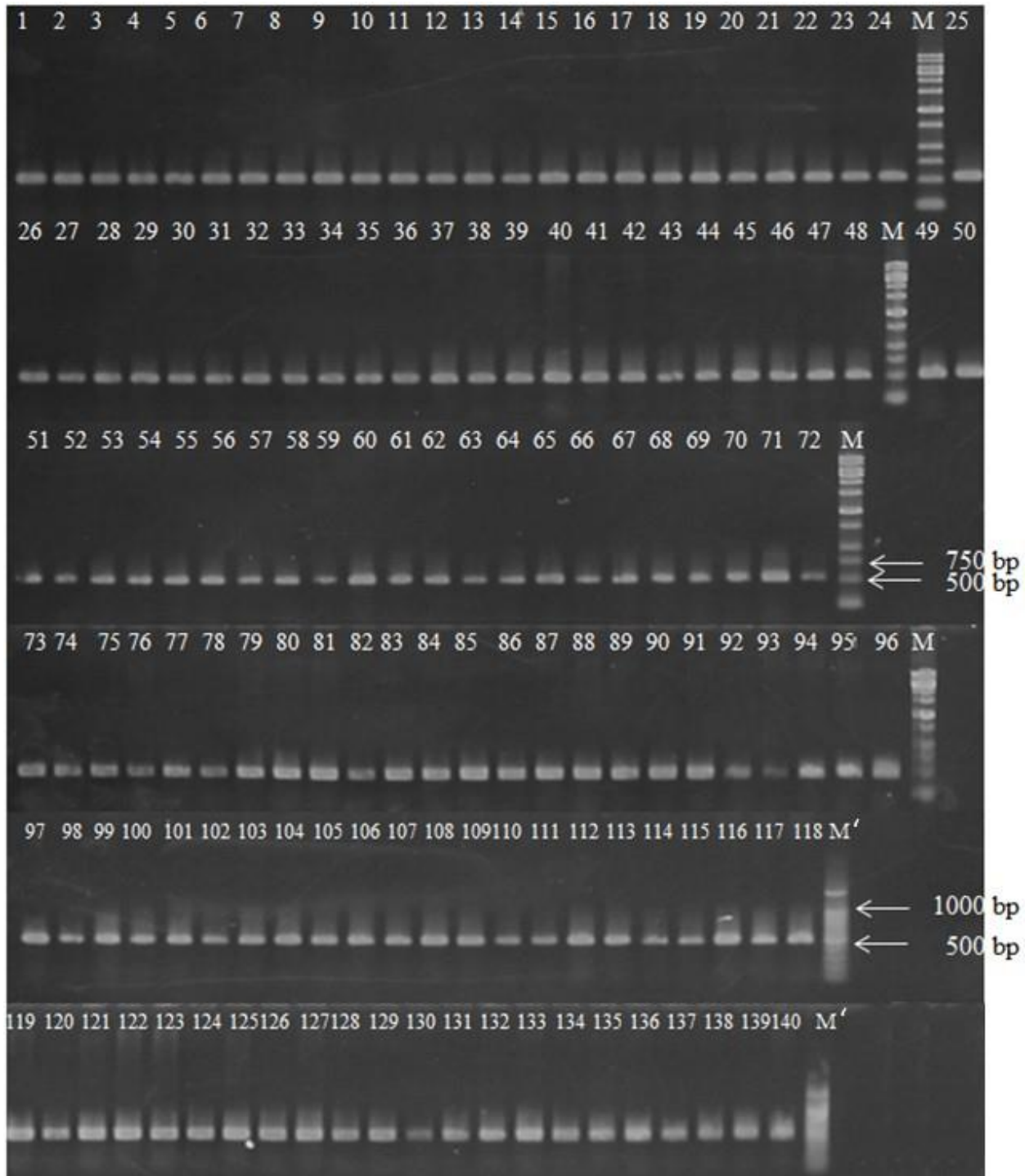
Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 501 bp chứa SNP *IGF1R*.15392 T>A từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



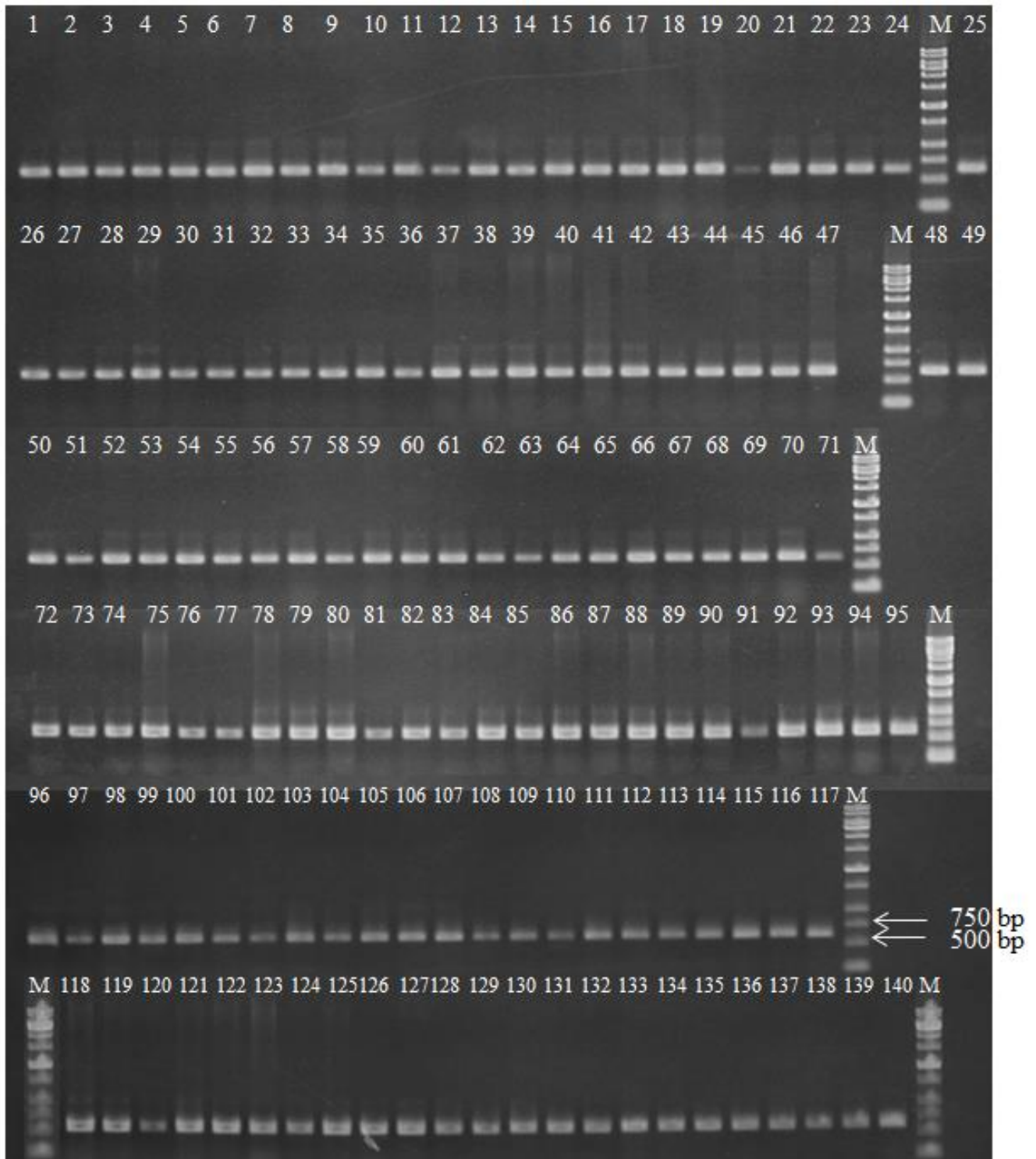
Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 586 bp chứa SNP *IGF1R*.83894 A>G từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



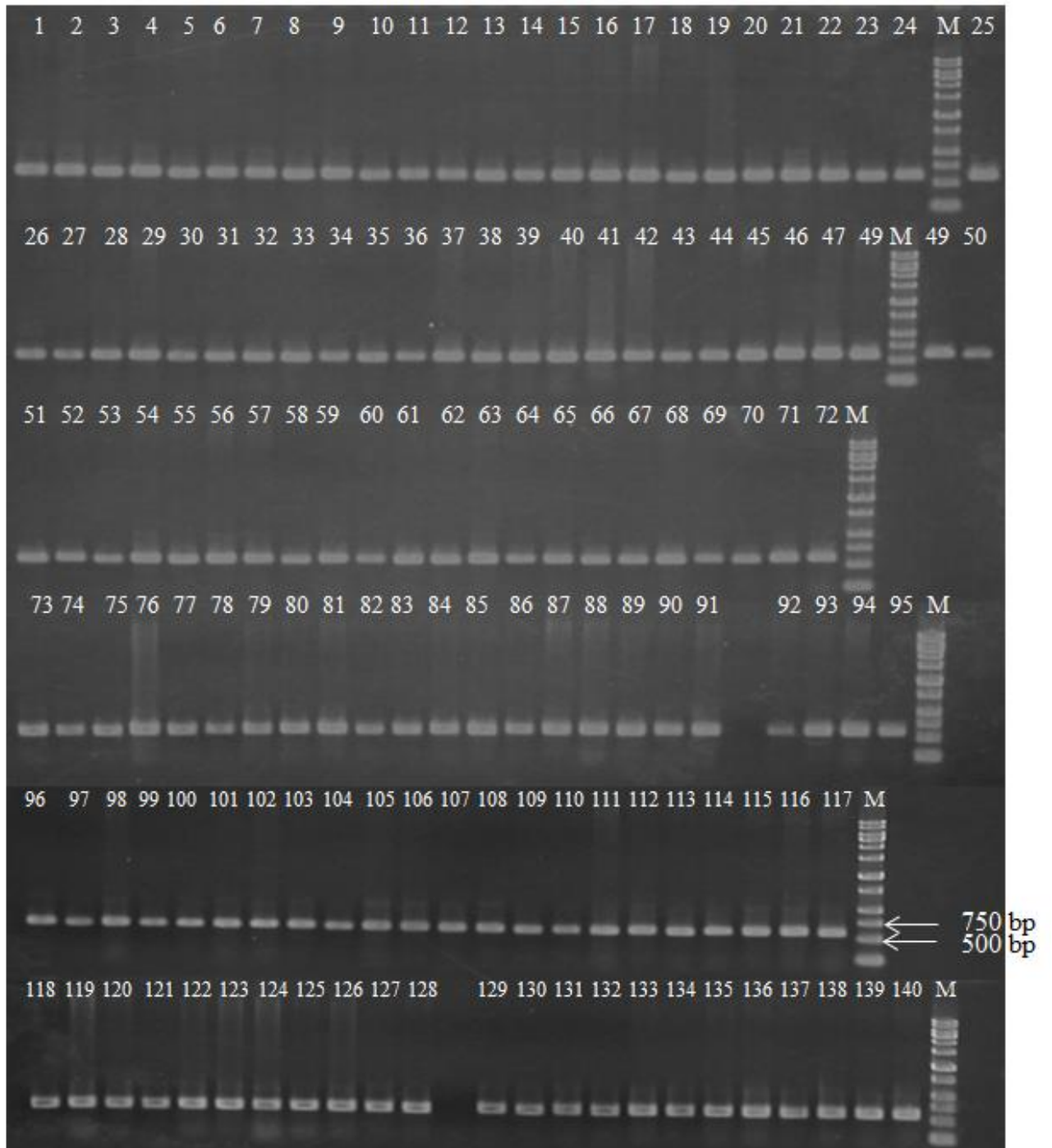
Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 553 bp chứa SNP *IGFBP-3.704 C>G* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



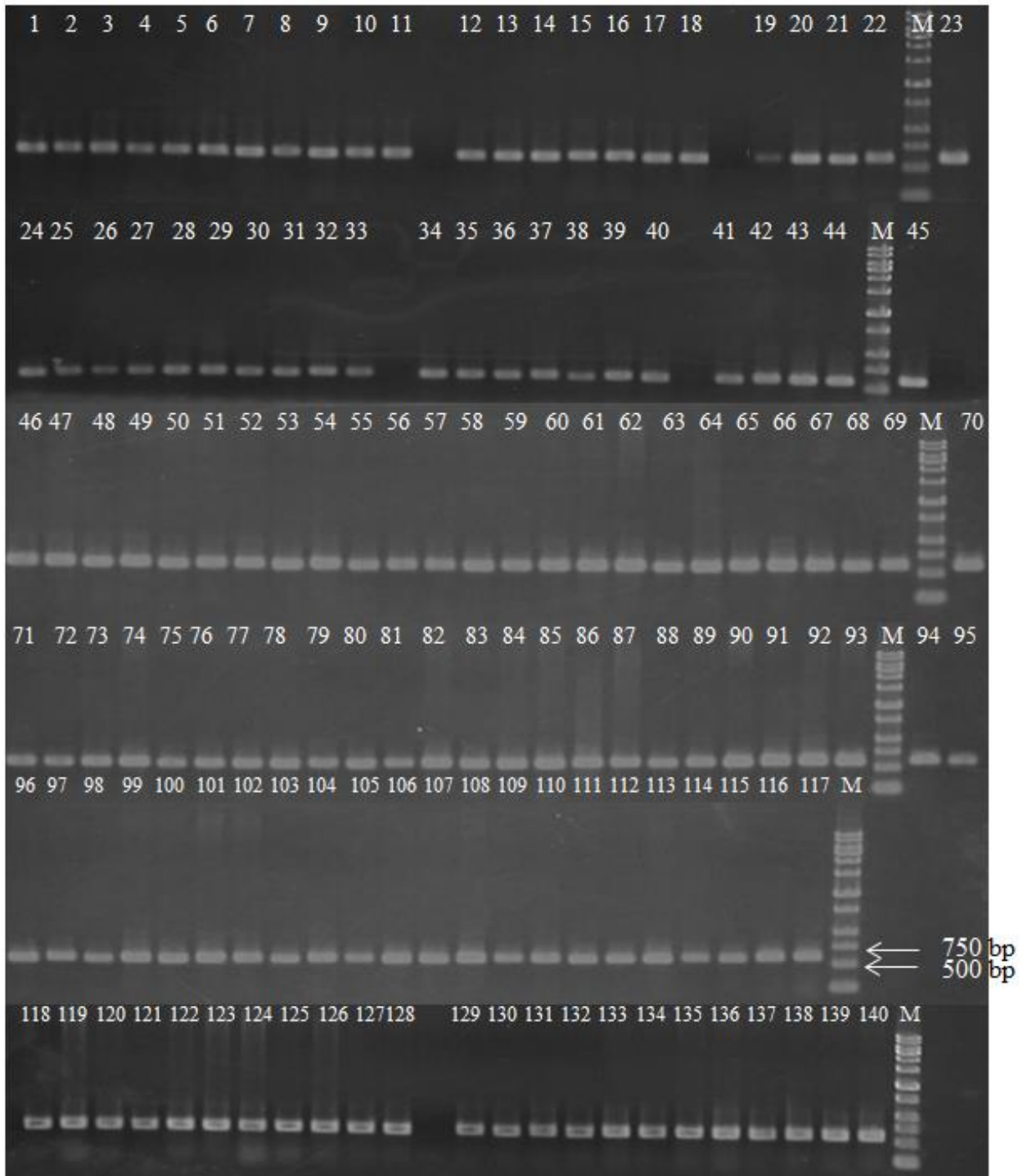
Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 547 bp chứa SNP *IGFBP-5.525 T>A* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



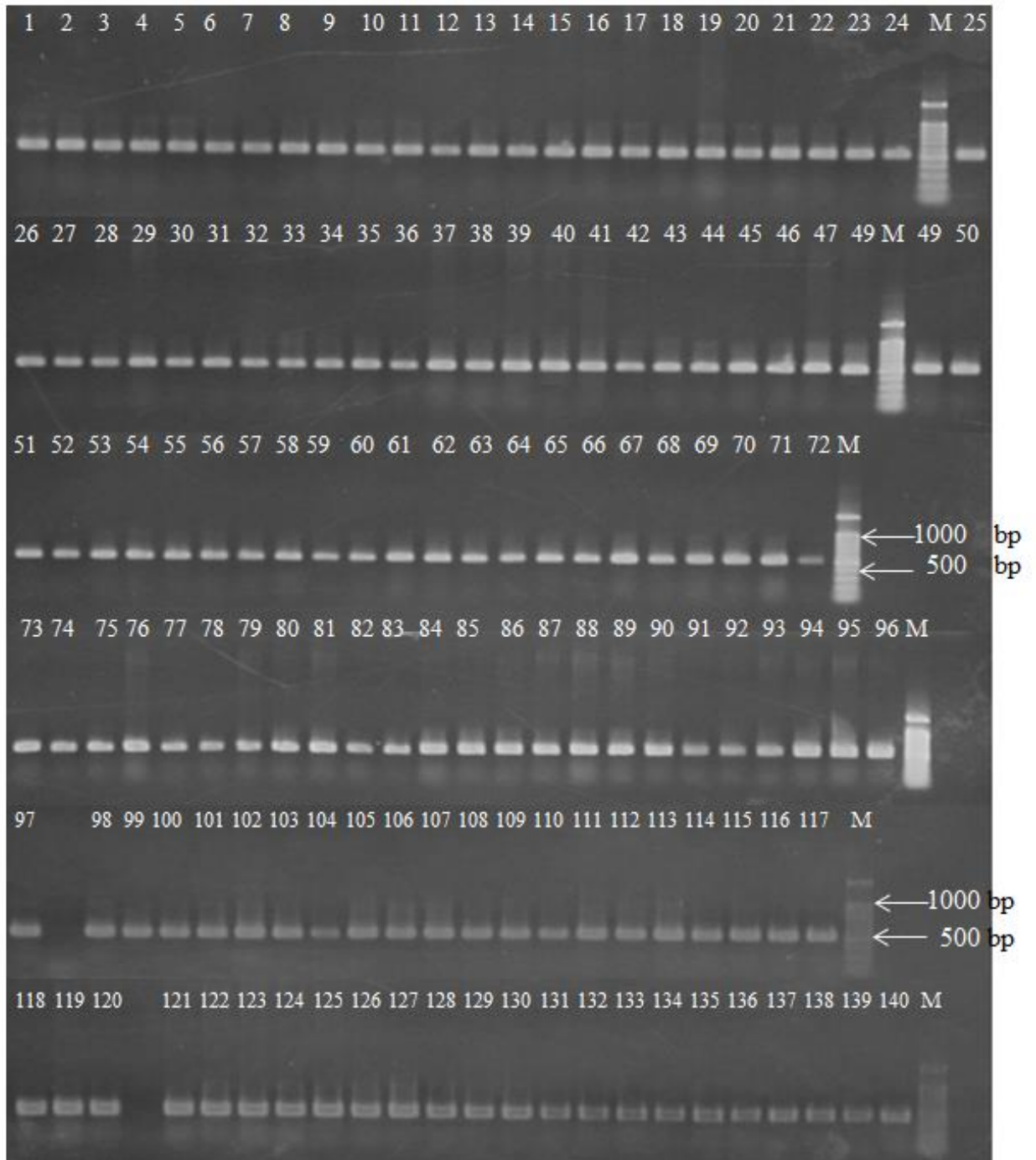
Kết quả điện di sản phẩm PCR khuếch đại từ cặp môi nhân đoạn dài 627 bp chứa SNP *IGFBP-6.2278 C>A* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



Kết quả điện di sản phẩm PCR khuếch đại từ cặp mồi nhân đoạn dài 518 bp chứa SNP *IGFBP-7.344 T>C* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



Kết quả điện di sản phẩm PCR khuếch đại từ cặp môi nhân đoạn dài 551 bp chứa SNP *IGFBP-7.2060 A>G* từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.



Kết quả điện di sản phẩm PCR khuếch đại từ cặp môi nhân đoạn dài 529 bp chứa SNP *IGFBP-7*. 4559 C>A từ DNA tổng số của 140 cá thể thuộc bộ mẫu kiểm nghiệm.

Phụ lục 9: Kết quả genotyping 10 SNP được kiểm nghiệm trên 160 cá thể sử dụng trong luận án.

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
1	TTN	N1	A	A	T	T	T	A	A	A	G	C	T	T	C	C	C	C	G	G	C	C
2	TTN	N2	A	A	T	T	T	A	G	G	G	G	T	T	C	C	C	C	G	G	A	A
3	TTN	N3	A	A	T	C	T	T	A	G	G	G	T	A	C	C	C	C	G	G	A	A
4	TTN	N4	A	A	T	C	T	A	A	A	G	G	T	A	C	C	C	C	G	G	A	A
5	TTN	N5	A	A	T	T	T	T	A	A	G	G	T	A	C	C	C	C	G	G	C	C
6	TTN	N6	A	A	T	C	T	A	G	G	G	G	T	T	A	A	C	T	G	G	A	C
7	TTN	N7	A	A	T	C	T	T	A	A	G	G	T	T	C	C	C	T	G	G	A	C
8	TTN	N8	T	T	T	C	T	T	G	G	G	G	T	T	C	C	C	C	G	G	A	C
9	TTN	N9	A	A	T	T	T	T	G	G	G	G	T	T	C	C	C	C	G	G	A	A
10	TTN	N10	A	A	T	C	T	A	G	G	G	G	T	T	A	A	T	T	G	G	A	A
11	TTN	3	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	C	G	G	A	A
12	TTN	41	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	C	G	A	A	A
13	TTN	122	T	A	C	T	0	0	G	A	G	G	A	T	A	A	C	T	G	A	A	A
14	TTN	142	A	T	T	T	0	0	G	A	G	G	0	0	A	A	C	C	G	A	A	A
15	TTN	177	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	C
16	TTN	181	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
			T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	C	G	A	A	A
17	TTN	200	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	C	G	A	A	A
18	TTN	2	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	C	G	G	A	A
19	TTN	42	A	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
20	TTN	118	T	A	C	T	T	A	G	A	G	C	A	T	A	A	C	C	G	A	A	A
21	TTN	123	T	A	C	T	0	0	G	A	G	G	A	T	C	A	C	T	G	A	A	A
22	TTN	128	A	A	T	T	T	A	G	A	G	C	A	T	C	A	T	C	G	A	A	A
23	TTN	159	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
24	TTN	195	T	A	C	T	T	A	G	A	G	G	A	T	A	A	C	T	G	A	A	A
25	TTN	205	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
26	TTN	13	A	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
27	TTN	23	T	A	C	T	T	A	G	A	G	G	A	A	C	A	C	C	G	A	C	C
28	TTN	24	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	A	A	A	A
29	TTN	31	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
30	TTN	35	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
31	TTN	38	T	A	C	T	0	0	G	A	G	G	A	T	C	A	C	C	G	G	A	A
32	TTN	57	T	A	C	T	T	A	A	A	G	C	A	T	C	A	C	T	G	A	A	A
33	TTN	59	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	C	G	A	C	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
			T	A	C	T	T	A	A	A	G	C	A	T	C	A	C	T	G	A	A	A
34	TTN	78	T	A	C	T	T	A	A	A	G	C	A	T	C	A	C	T	G	A	A	A
35	TTN	92	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
36	TTN	145	T	A	C	T	A	T	G	A	G	C	A	T	C	A	C	T	G	A	A	A
37	TTN	165	A	A	C	T	T	A	G	A	G	C	A	T	A	A	C	T	G	A	A	A
38	TTN	175	A	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
39	TTN	183	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	G	A	A
40	TTN	216	A	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
41	TTN	218	T	A	C	T	T	A	G	A	G	C	A	T	A	A	C	T	G	A	A	A
42	TTN	18	A	T	C	T	0	0	G	A	G	G	A	T	A	A	C	C	G	A	C	C
43	TTN	21	T	A	C	T	T	A	G	A	G	G	A	T	A	A	C	C	G	A	A	A
44	TTN	66	T	A	T	T	T	T	G	A	G	C	A	T	A	A	C	T	G	A	A	A
45	TTN	74	T	A	C	T	T	A	G	A	G	G	A	A	A	A	C	T	G	A	A	A
46	TTN	97	A	T	T	T	T	A	G	A	G	G	A	T	C	A	C	C	G	G	A	A
47	TTN	197	A	T	C	T	0	0	G	A	G	G	A	T	A	A	C	C	G	A	A	A
48	TTN	201	T	A	C	T	T	A	G	A	G	C	A	T	C	A	T	C	G	G	A	A
49	TTN	213	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
50	TTN	12	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
			T	A	C	T	T	A	G	A	G	C	A	T	C	A	T	C	T	C	G	A
51	TTN	30	T	A	C	T	T	A	G	A	G	C	A	T	C	A	T	C	G	A	A	A
52	TTN	43	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
53	TTN	63	T	A	C	T	T	A	A	A	G	C	A	A	C	A	C	C	G	A	A	A
54	TTN	68	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
55	TTN	90	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
56	TTN	102	T	A	C	C	T	A	G	A	G	C	A	T	C	A	T	C	G	A	C	A
57	TTN	107	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
58	TTN	109	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
59	TTN	11	A	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
60	TTN	20	T	A	C	T	T	A	G	A	G	C	A	T	C	A	T	C	G	A	A	A
61	TTN	83	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
62	TTN	88	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
63	TTN	94	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	G	A	A
64	TTN	121	A	T	T	T	T	A	A	A	G	C	A	T	C	A	C	T	G	G	A	A
65	TTN	130	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
66	TTN	134	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
67	TTN	158	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	C	G	A	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
			T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
68	TTN	167	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
69	TTN	168	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
70	TTN	194	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	G	A	A
71	TTN	411	T	A	T	T	T	A	G	A	G	G	A	T	A	A	C	C	G	A	A	A
72	TTN	412	T	A	T	C	T	A	G	A	G	C	A	T	C	A	C	T	G	G	C	A
73	TTN	157	A	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
74	TTN	166	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
75	TTN	170	A	T	T	T	T	A	A	A	G	C	A	T	C	A	C	T	G	G	A	A
76	TTN	172	A	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	G	A	A
77	TTN	204	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	G	A	A
78	TTN	101	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
79	TTN	143	T	A	T	T	A	T	G	A	G	C	A	T	C	A	C	T	G	A	A	A
80	TTN	151	T	A	C	T	T	A	G	A	G	G	A	T	A	A	C	T	G	A	A	A
81	TTC	16	A	A	C	C	T	A	A	A	G	G	T	T	C	A	C	T	A	A	A	A
82	TTC	40	T	A	T	T	A	A	G	A	G	G	T	T	C	A	C	C	G	A	A	A
83	TTC	55	T	A	C	C	T	A	G	A	G	G	T	T	C	A	C	T	A	A	C	A
84	TTC	67	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
85	TTC	72	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	T	A	A	C	A
86	TTC	75	T	A	T	T	T	A	G	A	G	G	A	T	A	A	C	C	A	A	A	A
87	TTC	120	T	A	T	T	0	0	G	A	G	G	T	T	C	A	C	T	A	A	A	A
88	TTC	179	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	C	A
89	TTC	186	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	T	G	A	A	A
90	TTC	188	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	C	A
91	TTC	14	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
92	TTC	49	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	T	A	A	A	A
93	TTC	53	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	G	A	A	C
94	TTC	61	T	A	C	T	A	T	G	A	G	C	A	T	C	A	C	T	A	A	C	A
95	TTC	69	A	T	T	T	T	A	G	A	G	G	T	T	C	A	C	T	A	A	C	A
96	TTC	77	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	G	A	A	A
97	TTC	82	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	A	A
98	TTC	84	T	A	C	C	T	A	G	A	G	C	A	T	A	A	C	T	G	A	A	A
99	TTC	135	T	A	C	C	T	A	G	A	G	C	T	T	C	A	C	T	G	A	A	A
100	TTC	155	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	T	G	A	A	A
101	TTC	182	0	0	C	C	0	0	G	A	G	C	T	T	C	A	C	T	A	A	A	A

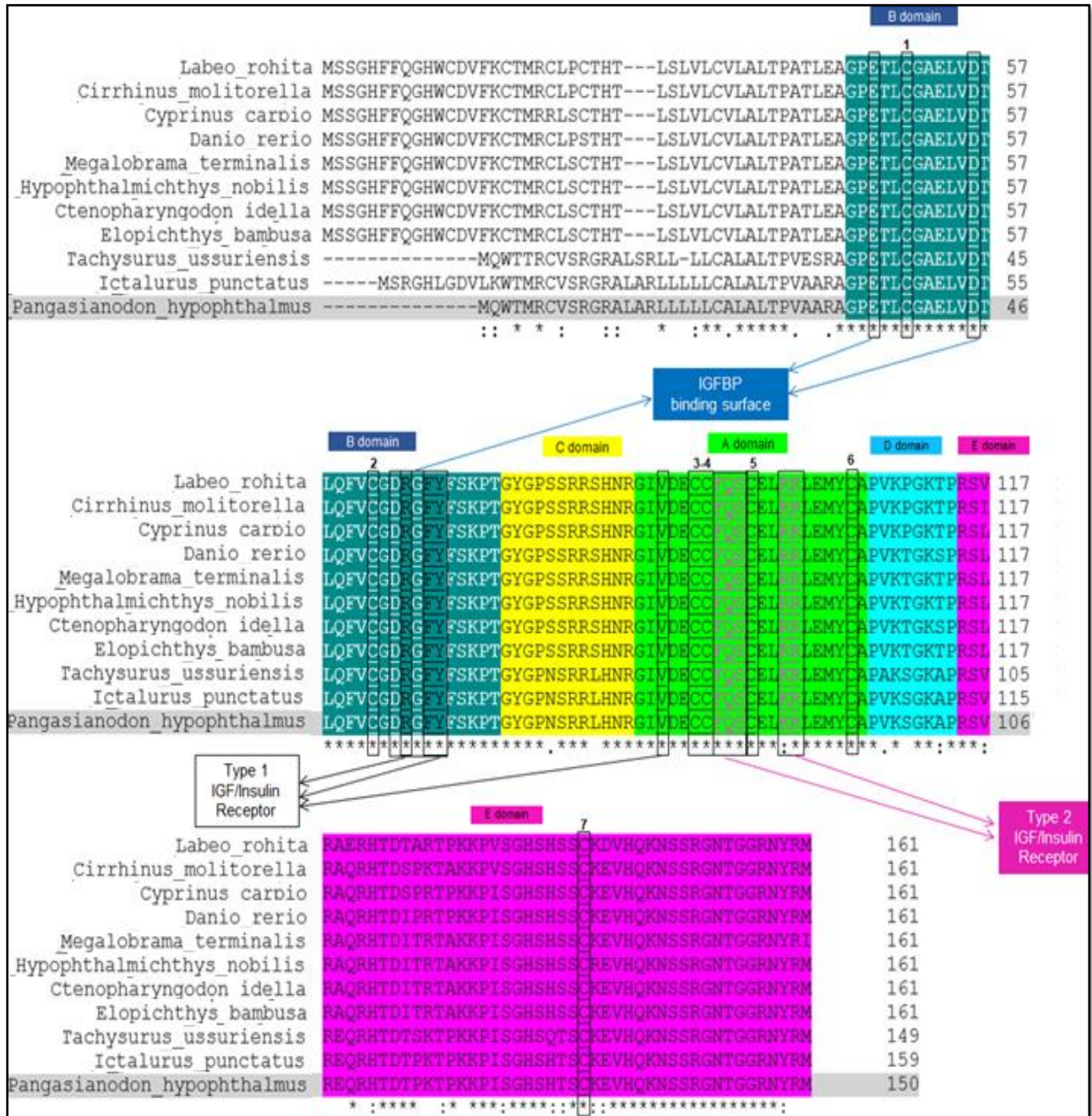
STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
			T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	C	A	A	C	A
102	TTC	189	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	C	A	A	C	A
103	TTC	191	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	C	A
104	TTC	207	T	T	C	T	0	0	G	A	G	G	T	T	C	C	C	C	G	G	C	A
105	TTC	209	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	C	A
106	TTC	212	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	T	G	A	A	A
107	TTC	215	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
108	TTC	7	T	A	C	T	T	A	G	A	G	G	A	T	A	A	C	C	G	A	A	A
109	TTC	45	A	A	C	T	T	A	G	A	G	G	T	T	C	A	C	C	G	A	A	A
110	TTC	47	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	C	G	A	A	A
111	TTC	111	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
112	TTC	138	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	C
113	TTC	148	T	A	T	T	T	A	G	A	G	C	A	T	A	A	C	C	A	A	A	A
114	TTC	162	T	A	C	C	T	A	G	A	G	G	T	T	C	A	C	C	G	A	A	A
115	TTC	173	A	A	C	T	T	A	G	A	G	G	A	T	A	A	C	C	A	A	C	C
116	TTC	199	T	A	C	T	T	A	G	A	G	G	0	0	0	0	0	0	0	0	0	0
117	TTC	9	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	C	A	A	A	A
118	TTC	146	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
119	TTC	149	T	A	T	T	T	A	G	A	G	G	A	T	A	A	C	C	G	A	A	A
120	TTC	150	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	C	A	A	C	C
121	TTC	202	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	C	G	A	A	A
122	TTC	46	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	T	G	A	C	A
123	TTC	54	T	A	T	T	T	A	G	A	G	G	T	T	C	A	C	C	G	A	C	A
124	TTC	76	T	A	C	T	T	A	G	A	0	0	0	0	C	C	0	0	0	0	0	0
125	TTC	79	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
126	TTC	80	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
127	TTC	125	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
128	TTC	133	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
129	TTC	136	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
130	TTC	137	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
131	TTC	34	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
132	TTC	96	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
133	TTC	131	T	A	C	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
134	TTC	141	A	T	C	T	T	A	G	A	G	G	T	T	A	A	C	T	A	A	A	A
135	TTC	180	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	C

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
136	TTC	190	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
137	TTC	192	A	T	C	T	T	A	G	A	G	C	A	T	A	A	C	T	A	A	A	A
138	TTC	418	0	0	T	T	0	0	G	A	G	G	T	T	C	C	C	T	G	A	A	A
139	TTC	62	T	A	T	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	C	A
140	TTC	112	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	A	A	A	A
141	TTC	144	T	A	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
142	TTC	17	T	A	C	T	T	A	G	A	G	G	T	T	A	A	C	C	A	A	A	A
143	TTC	52	T	A	C	C	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
144	TTC	119	A	T	C	T	T	A	G	A	G	C	A	T	C	A	C	T	G	A	A	A
145	TTC	185	T	A	C	T	T	A	G	A	G	G	T	T	C	A	C	C	G	A	A	A
146	TTC	409	T	A	C	T	T	A	G	A	G	C	A	T	A	A	C	T	A	A	A	A
147	TTC	124	T	A	C	T	T	A	G	A	G	G	T	T	A	A	C	C	G	A	C	A
148	TTC	176	T	T	0	0	0	0	G	A	G	G	T	T	C	A	C	C	A	A	C	A
149	TTC	211	A	T	C	T	T	A	G	A	G	G	T	T	C	A	C	C	A	A	C	C
150	TTC	220	T	A	T	T	T	A	G	A	G	G	A	T	C	A	C	T	G	A	A	A
151	TTC	C1	A	A	T	C	T	T	G	G	G	G	T	T	C	C	T	T	G	G	A	A
152	TTC	C2	T	T	T	C	T	T	G	G	G	G	T	T	A	A	T	T	G	G	A	A

STT	Loại	Kí hiệu mẫu	<i>IGF1</i> 13680 A>T		<i>IGF1R</i> 13357 T>C		<i>IGF1R</i> 15392 T>A		<i>IGF1R</i> 83894 A>G		<i>IGFBP3</i> 704 C>G		<i>IGFBP5</i> 525 T>A		<i>IGFBP6</i> 2278 C>A		<i>IGFBP7</i> 344 T>C		<i>IGFBP7</i> 2060 A>G		<i>IGFBP7</i> 4559 C>A	
153	TTC	C3	A	T	T	C	T	T	G	G	G	G	T	T	C	C	C	C	G	G	A	A
154	TTC	C4	T	T	T	C	T	T	G	G	G	G	T	T	A	A	T	T	G	G	A	A
155	TTC	C5	A	A	C	C	T	T	G	G	G	G	T	T	A	A	C	C	A	A	A	A
156	TTC	C6	T	T	T	C	T	T	G	G	G	C	T	T	C	C	C	T	A	A	A	A
157	TTC	C7	T	T	T	C	T	T	G	G	G	G	T	T	A	A	C	T	G	G	A	A
158	TTC	C8	A	A	T	C	T	T	G	G	G	G	T	T	A	A	C	C	G	A	A	A
159	TTC	C9	T	T	C	C	T	T	G	G	G	C	T	T	C	C	C	C	G	A	A	A
160	TTC	C10	A	A	T	C	T	T	A	A	G	G	T	T	A	A	C	T	A	A	A	A

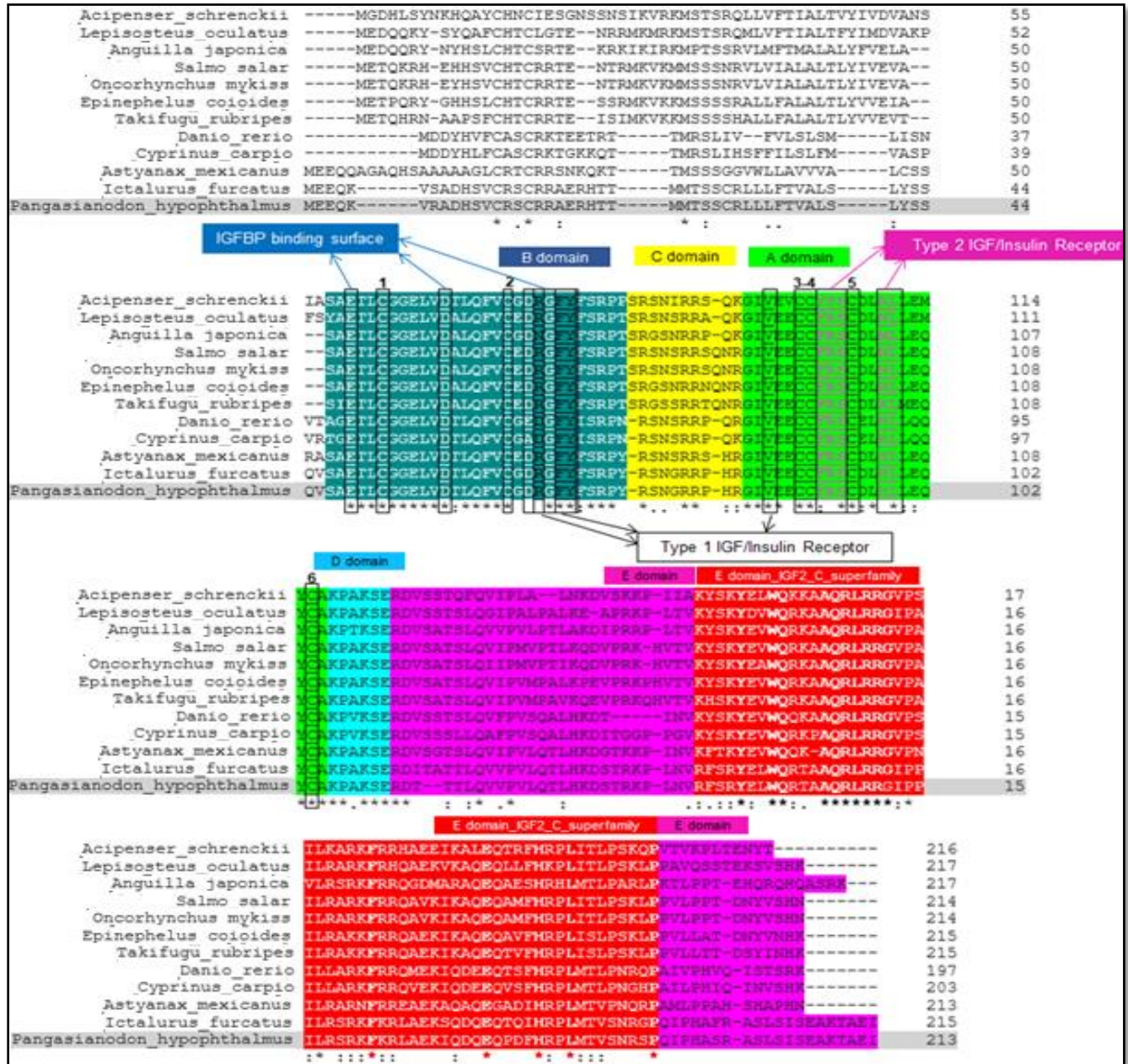
Phụ lục 10: So sánh trình tự protein IGF1 của cá tra nuôi và các loài cá khác.



Labeo rohita: cá trôi Ấn Độ, *Cirrhinus molitorella*: cá trôi đen, *Cyprinus carpio*: cá chép, *Danio rerio*: cá ngựa vằn, *Megalobrama terminalis*: cá bống đen Amur, *Hypophthalmichthys nobilis*: cá mè hoa, *Ctenopharyngodon idella*: cá trắm cỏ, *Elopichthys bambusa*: cá vàng (kanyu), *Tachysurus ussuriensis*: cá da trơn Ussuri, *Ictalurus punctatus*: cá nheo Mỹ.

Các khung viền đen đánh số 1,2,3,4,5,6,7: 7 vị trí Cystein (C) bảo tồn. Các trình tự axit amin đóng khung: vị trí bám với các protein chức năng.

Phụ lục 11: So sánh trình tự protein IGF2 của cá tra nuôi và các loài cá khác.



Acipenser schrenckii: Cá tầm sông Amur/Cá hoàng, *Lepisosteus oculatus*: Cá lóc đốm, *Anguilla japonica*: Cá chình Nhật Bản, *Salmo salar*: Cá hồi Đại Tây Dương, *Oncorhynchus mykiss*: Cá hồi cầu vồng/Cá hồi vân, *Epinephelus coioides*: Cá mú đen chấm nâu, *Takifugu rubripes*: Cá nóc hổ, *Danio rerio*: Cá ngựa vằn, *Cyprinus carpio*: Cá chép, *Astyanax mexicanus*: Cá hang mù Mexico, *Ictalurus furcatus*: Cá nheo lục

Các khung viền đen đánh số 1,2,3,4,5,6: 6 vị trí Cystein (C) bảo tồn. Các trình tự axit amin đồng khung: vị trí tám với các protein chức năng.

XP_027017178.1 Tachysurus_fulvidraco	LPIGVLLIFVTISTFFLVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ¹⁰¹⁶	1016	Motif gắn với cơ chất của thụ thể insulin (997 đến 1000)
KAG7333206.1 Hemibagrus_wyckioides	LPIVLLILVFLIFTIIIFVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ¹⁰¹⁵	975	
KAF7699092.1 Silurus_meridionalis	LPIIVLLIFITLIVTIIIFVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ¹⁰¹⁵	1015	
XP_026783005.2 Pangasianodon_hypophthalmus	LPVIVLLLVLTIIITIIIFVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ¹⁰¹⁹	1019	
XP_017321963.1 Ictalurus_punctatus	LPIIVLLILATLITIIIFVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ⁹⁹⁹	999	
KAF4091542.1 Ameiurus_melas	LPIIVLLILATLITIIIFVKKRNSDRLGNGVLYASVNP ⁹⁷⁵ YFSAAE ⁹⁷⁶ MYVPDE ⁹⁷⁷ WEVAREK ⁹⁹⁸	998	
XP_027017178.1 Tachysurus_fulvidraco	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰⁷⁶	1076	Vị trí bám của ATP (1026 đến 1055)
KAG7333206.1 Hemibagrus_wyckioides	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰³⁵	1035	
KAF7699092.1 Silurus_meridionalis	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰⁷⁵	1075	
XP_026783005.2 Pangasianodon_hypophthalmus	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰⁷⁹	1079	
XP_017321963.1 Ictalurus_punctatus	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰⁵⁹	1059	
KAF4091542.1 Ameiurus_melas	TLCREL ¹⁰²⁶ CGSFG ¹⁰²⁷ WVYEGIAKGV ¹⁰²⁸ WDEPETRVAIK ¹⁰²⁹ TVNESASMRERIEFLNEASVMKEFNC ¹⁰⁵⁸	1058	
XP_027017178.1 Tachysurus_fulvidraco	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹¹³⁶	1136	Domain PTKc_InSR_like (1012 đến 1288)
KAG7333206.1 Hemibagrus_wyckioides	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹⁰⁹⁴	1094	
KAF7699092.1 Silurus_meridionalis	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹¹³⁵	1135	
XP_026783005.2 Pangasianodon_hypophthalmus	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹¹³⁹	1139	
XP_017321963.1 Ictalurus_punctatus	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹¹¹⁹	1119	
KAF4091542.1 Ameiurus_melas	HHVVRLLGVVSGQ ¹¹¹⁸ QPTLVIMELMTRGDLKSYLRSLRSKQ ¹¹¹⁹ QSSCSQALPPLK ¹¹²⁰ KMIQ ¹¹²¹ MAGEI ¹¹¹⁸	1118	
XP_027017178.1 Tachysurus_fulvidraco	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁹⁶	1196	Vị trí phosphoryl hóa tự động (1181 đến 1186)
KAG7333206.1 Hemibagrus_wyckioides	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁵⁴	1154	
KAF7699092.1 Silurus_meridionalis	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁹⁵	1195	
XP_026783005.2 Pangasianodon_hypophthalmus	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁹⁹	1199	
XP_017321963.1 Ictalurus_punctatus	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁷⁹	1179	
KAF4091542.1 Ameiurus_melas	SDGMAYLNANKEFVHRDLAARNCMVAEDFTVKIGDFGMTRDI ¹¹⁷⁸ YETDYY ¹¹⁷⁹ RRKGGKGLLPV ¹¹⁸⁰ RRM ¹¹⁷⁸	1178	
XP_027017178.1 Tachysurus_fulvidraco	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²⁵⁶ PD	1256	
KAG7333206.1 Hemibagrus_wyckioides	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²¹⁴ PD	1214	
KAF7699092.1 Silurus_meridionalis	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²⁵⁵ PD	1255	
XP_026783005.2 Pangasianodon_hypophthalmus	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²⁵⁹ PD	1259	
XP_017321963.1 Ictalurus_punctatus	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²³⁹ PD	1239	
KAF4091542.1 Ameiurus_melas	SPESLKDGVFTTNSDVNSFGV ¹²¹⁴ LWEIATLAEQPYQGSNEQVLR ¹²¹⁵ RFVMEGGLLDK ¹²¹⁶ PDNC ¹²³⁸ PD	1238	
XP_027017178.1 Tachysurus_fulvidraco	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFHDVSVFFYSDENKPPDTEELDVEEN ¹³¹⁴	1314	
KAG7333206.1 Hemibagrus_wyckioides	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFRDVSFFYSDENKPPDTEELDVEEN ¹²⁷⁴	1274	
KAF7699092.1 Silurus_meridionalis	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFSEVSVFFYSDENKPPDTEELDVEEN ¹³¹⁵	1315	
XP_026783005.2 Pangasianodon_hypophthalmus	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFRDVSFFYSDENKPPDTEELDVEEN ¹³¹⁹	1319	
XP_017321963.1 Ictalurus_punctatus	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFHDVSVFFYSDENKPPDTEELDVEEN ¹²⁹⁹	1299	
KAF4091542.1 Ameiurus_melas	MLFELMRMCWQYNPKMRPSFLEIISSIKE ¹³¹² LEPSFHDVSVFFYSDENKPPDTEELDVEEN ¹²⁹⁸	1298	
XP_027017178.1 Tachysurus_fulvidraco	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³⁷⁴ DNTQ	1374	
KAG7333206.1 Hemibagrus_wyckioides	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³³⁴ DNTQ	1334	
KAF7699092.1 Silurus_meridionalis	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³⁷⁵ DNTQ	1375	
XP_026783005.2 Pangasianodon_hypophthalmus	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³⁷⁹ DNTQ	1379	
XP_017321963.1 Ictalurus_punctatus	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³⁵⁹ DNTQ	1359	
KAF4091542.1 Ameiurus_melas	MENIPLDPASTLT ¹³⁵⁸ PPSSQGLTETATLPSAPMCTSVPTPSTSPSSSTASPALDSHF ¹³⁵⁸ DNTQ	1358	
XP_027017178.1 Tachysurus_fulvidraco	PYAHMNGGRKNERALPLPQSSAC 1397	1397	
KAG7333206.1 Hemibagrus_wyckioides	PYAHMNGGRKNERALPLPQSSAC 1357	1357	
KAF7699092.1 Silurus_meridionalis	PYAHMNGGRKNERALPLPQSSAC 1398	1398	
XP_026783005.2 Pangasianodon_hypophthalmus	PYAHMNGGRKNERALPLPQSSAC 1402	1402	
XP_017321963.1 Ictalurus_punctatus	PYAHMNGGRKNERALPLPQSSAC 1382	1382	
KAF4091542.1 Ameiurus_melas	PYAHMNGGRKNERALPLPQSSAC 1381	1381	

Tachysurus fulvidraco: cá trê đầu vàng, *Hemibagrus wyckioides*: cá lăng nha, *Silurus meridionalis*: cá trê đá Biwa, *Ictalurus punctatus*: cá nheo Mỹ, *Ameiurus melas*: cá da tron đen.