MINISTRY OF EDUCATION AND TRAINING

VIETNAM ACADEMY OF SCIENCE AND TECHNOLOGY

GRADUATE UNIVERSITY OF SCIENCE AND TECHNOLOGY

TRAN THI HUYEN TRANG

CHARACTERIZATION OF SINGLE NUCLEOTIDE POLYMORPHISMS IN SOME GENES ASSOCIATED WITH GROWTH TRAITS IN STRIPED CATFISH (*Pangasianodon hypophthalmus*)

Major: Biotechnology
Code: 9 42 02 01

SUMMARY OF BIOTECHNOLOGY DOCTORAL THESIS

Hanoi – 2023

The thesis was completed at: Graduate University of Science and Technology – Vietnam Academy of Science and Technology
Institute of Genome Research

Advisor: Dr. Kim Thi Phuong Oanh

Reviewer 1: Assos.Prof.Dr. Ngo Thi Kim Cuc
Reviewer 2: Assos.Prof.Dr. Phan Ke Long
Reviewer 3: Assos.Prof.Dr. Nguyen Thi Hong Van

The thesis was defended at the doctoral thesis review committee, meeting in Graduate University of Science and Technology, Vietnam Academy of Science and Technology from 9 am to 11:30 am, June 27th 2023

The thesis can be found in:
- Library of Graduate University of Science and Technology
- National Library of Vietnam

INTRODUCTION

1. The necessary of thesis

The striped catfish, *Pangasianodon hypophthalmus* is naturally distributed in the Chao Phraya river of Thailand and the Mekong river, which flows through Cambodia, Laos, Thailand, and Vietnam. According to Prime Minister's decision No. 50/2018/QĐ-TTg, *P. hypophthalmus* is the major fish species cultured in Vietnam for export and domestic consumption. However, striped catfish's farming in Vietnam has faced to many challenges such as the reducing quality of brood stock and the annual endemic causing economic loss to farmers. To enhance the pangasius production efficiency with long-term sustainability, improvement of brood stocks is one of the most important issues. In which, finding marker assisted selection (MAS), especially polymorphisms in candidate genes, associated to growth traits is effective approach. The Insulin-like Growth Factor (IGF) system includes IGF ligands (IGF1, IGF2), IGF receptors (IGFRs), and IGF-binding protein (IGFBPs), plays a central role in the neuroendocrine regulation of growth in all vertebrates. Therefore, the characterization of single nucleotide polymorphisms in some genes of IGF system associated with growth traits of striped catfish is very necessary, contributing to find marker assisted selection for growth traits of this fish and serving the demand of aquaculture in Vietnam.

2. The aims of thesis

This thesis aimed to discover SNPs in some genes of IGF system, characterize the association of these SNPs with the growth traits of striped catfish (*Pangasianodon hypophthalmus*), and suggest potential markers assisted selection for the growth of pangasius breeding.

3. The main contents of thesis

- Structure analysis of *IGF1, IGF2, IGF1R, IGFBP-1,-2 -3,-5,-6,-7* genes in IGF system of striped catfish, finding the important fragments to characterize SNPs, Sanger sequencing these fragments to evaluate the authenticity of Sanger sequences and the reference sequences.

- Sanger sequencing the important fragments in 9 above mentioned genes in the discovery sample set including 10 fast- growing fish and 10 slow- growing fish, aligning the corresponding sequences and the

reference sequence to detect SNPs, filtering these SNPs by criteria.

- Individual genotyping filtered SNPs in the validation sample set including 70 fast- growing fish and 70 slow-growing fish by Single base extension (SBE) method.

- Analysis on SNP data collected from discovery and validation sample sets including 80 fast- growing fish and 80 slow- growing fish, characterizing the association of filtered SNPs, haplotypes, diplotypes with the growth traits of striped catfish.

CHAPTER 1. OVERVIEWS

1.1. The introduction of striped catfish *Pangasianodon hypophthalmus*

The striped catfish, *Pangasianodon hypophthalmus* (Sauvage, 1878) with the initial name as *Helicophagus hypophthalmus* (Sauvage, 1878) and other name as *Pangasius sutchi*, classified in *Pangasius* genius, shark catfish family (*Pangasiidae*), catfish order (*Siluriformes*), is naturally distributed in the Chao Phraya river of Thailand and the Mekong river flowing through Cambodia, Laos, Thailand, and Vietnam. According to Prime Minister's decision No. 50/2018/QĐ-TTg, *P. hypophthalmus* is the major fish species cultured in Vietnam for export and domestic consumption. In 2022, pangasius industry expects export turnover to reach over 1.6 billion USD. To achieve this target, the Ministry of Agriculture and Rural Development requires localities to improve the quality of pangasius breeds, direct the production and supply enough high-quality brood stocks to improve production efficiency and reduce loss and lower production costs. In 2018, the draft genome of striped catfish was sequenced by NGS with the nuclear genome estimated at approximately 700 Mb, containing 28,600 protein-coding genes. It can be used as reference genome to develop molecular markers associated with traits of interest in breeding selection.

1.2. The growth (muscle growth) in fish and candidate genes

Muscle growth is one of the most important growth processes in fish, which is regulated by the neuroendocrine axis partly including the Insulin-like Growth Factor (IGF) system. Many previous studies showed the association between genetic variations in candidate genes of

neuroendocrine axis, especially in genes of IGF system, with the growth traits of teleost fish.

1.3. The function and structure of genes/proteins in IGF system

The Insulin-like Growth Factor (IGF)) includes IGF1, IGF2 ligands, IGF receptors (IGFRs) and IGF binding proteins (IGFBPs). The interaction of IGFs, IGFRs, IGFBPs and other proteins in signaling pathway induces the growth, differentiation and cell proliferation. IGF1 and IGF2 promotes the growth by enhancing the division and differentiation of skeletal muscle satellite cells, stimulating protein biosynthesis and development of aneurysms, and inhibiting proteolysis as well as muscle atrophy. IGF1R can bind and be activated by IGF1 and IGF2, while IGF2R is only activated by IGF2 leading to the degradation of this ligand. In the extracellular, IGFBPs carry IGFs, and modulate the bioavailability and the distribution of IGFs through the intervention to IGF – IGFR interaction. IGFBPs are secreted protein, and are classified into two groups, IGFBP family (IGFBP1-6) binding to IGF with high affinity and IGFBP- related proteins (IGFBP-rP 1-9) binding to IGF with low affinity. IGFBP-rP1 has other name as IGFBP-7, this protein has similar structure with IGFBP1-6.

While *IGF1* gene of mammal contains 6 exons, this gene of fish includes only 5 exons. The *IGF2* gene with 4 exons in fish is simple and more stable than that of mammal. Both of *IGF1Ra* and *IGF1Rb* genes in fish contain 21 exons, intercalated by introns with diverse size. All of *IGFBP*s in IGFBP family (IGFBP1-6) include 4 exons coding to their corresponding proteins, which contain IGFBP domain in N terminal, thyroglobulin type 1 domain in C terminal and linker domain in between.

1.4. The facility of characterizing the association between SNP in candidate genes of IGF system and the growth traits of striped catfish

The role of SNPs in determining molecular markers for traits of interest is proven in genomic as well as genetic studies. In aquaculture, SNPs are commonly applied in determining potential molecular markers associated with economic traits, such as growth and disease resistance. At the scale of finding SNPs in candidate genes, SNP methods based on PCR using probes or sequencing are widely applied. In which, the single base extension (SBE) method can detect tetra – allelic SNPs with high

sensitivity and specificity, and can automatically sequence in a short time. Many previous studies indicated the association between SNPs, haplotypes, combined genotypes in *IGF1*, *IGF2*, *IGF1R* genes of IGF systems and the growth of teleost fish, such as Atlantic salmon *Salmo salar*, common carp *Cyprinus carpio*, European sea bass *Dicentrarchus labrax*, Nile tilapia *Oreochromis niloticus* and freshwater sleeper native to China *Odontobutis potamophila*. Based on reference genome, 19 genes in IGF system of striped catfish were identified, including 2 genes encoding IGF1, 2 genes encoding IGF2, 4 genes encoding IGF1R and 11 genes encoding IGFBP-1,-2,-3,-5,-6,-7. All of these above facilitates contributed the experiments to characterize SNPs in candidate genes of IGF system associated with growth traits of striped catfish.

CHAPTER 2. MATERIALS AND METHODS

2.1. Materials

Individuals of striped catfish (*P. hypophthalmus*) sampled in this thesis were described in previous studies. In details, 160 individuals were selected from G3-merged families (2015) of growth selected line of striped catfish, which had been produced by a merit-based genetic selection at Research Institute of Aquaculture No.2. To discover SNPs, 10 fast-growing fish (top 10 highest estimated breeding value (EBV) individuals from 9 highest EBV families) and 10 slow growing fish (lowest EBV individuals from 9 lowest EBV families) were sampled into discovery sample set. In addition, 70 individuals with highest ranking of EBV from 24 highest EBV families and 70 individuals with lowest ranking of EBV from 31 lowest EBV families were sampled for SNP validation.

2.2. Methods

2.2.1. Structure analysis of genes

Based on the reference sequence of *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1,-2, -3,-5,-6,-7* genes of striped catfish, structure of genes and proteins was analyzed by BLASTn, ProteinBLAST (blastp), SignalP5.0 software to determine the important fragments of genes coding to functional domains, peptide signal and regulated regions in transcription/ translation. These fragments were sequenced in individuals to discover SNPs.

2.2.2. DNA extraction

Total DNA from fin clips was extracted by standard phenol/chloroform method. The quantity and quality of the extracted DNA were checked by NanoDrop One spectrophotometer (Thermo Fisher Scientific) and electrophoresis on 1% agarose gel.

2.2.3. Fragments of genes amplification

To amplify determined fragments of *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*,*-2*,*-3*,*-5*,*-6*,*-7* genes, polymerase chain reactions (PCR) were performed using total DNA as template and 55 primer pairs designed by Primer 3 (v.0.4.0) software. The quality of PCR products was check by electrophoresis on 1% agarose gel, and then the products were purified by Thermo Scientific GeneJET PCR Purification Kit (Thermo Fisher) according to the manufacturer's instructions.

2.2.4. Sanger sequencing

The purified PCR products were used as template for Sanger sequencing reaction using BigDye Terminator v3.1 Cycle Sequencing Kit (ThermoFisher Scientific) according to the manufacturer's instructions. After being purified, these products were analyzed by ABI®3500 Genetic Analyzer (Applied Biosystems). The ABI sequencer trace files were analyzed using BioEdit software and the presence of "double-peaks" at heterozygous sites was examined by eye.

2.2.5. Evaluation the similarity between Sanger sequence of fragments and reference genes

The Sanger sequence of fragments in an individual was compared with reference genes by BLASTn software to evaluate the similarity between them and reflecting the authenticity of Sanger sequence with reference sequence in genome sequenced by NGS.

2.2.6. Detecting, filtering SNPs in discovery sample set

To detect SNPs, the Sanger sequence of fragments in 20 individuals of discovery sample set was aligned to reference sequence to identify putative SNPs using MUSCLE software. SNP was filtered, if it was sequenced in total of 20 individuals (Non-identified genotype individual (NN) = 0), being non-synonymous SNP[(1)], or the SNP ratio in one group was at least 0.3[(2)], and their corresponding genotypic composition and/or allelic composition in two groups significantly differed[(3), (4)]. In details, the

SNP ratio of each group was calculated by the number of genotype of minor allele divided by the number of genotype of major allele. Minor allele had smaller ratio in total of 20 individuals than major allele had. The significant difference between fast- and slow-growing groups was assessed by using p-value from Fisher's exact test (p-value < 0.05), which was calculated by SHEsis software.

2.2.7. SNP validation by Single base extension (SBE)

The filtered SNPs were further validated by individually genotyping of 70 fast-growing fish and 70 slow-growing fish by SBE method. To reduce the amount of reaction, these SNPs were divided into two groups of multiplex SBE reaction performed by ABI SNaPshot Multiplex PCR Kit (Applied BioSystems), which could determine up to 10 SNPs per reaction. Based on the sequence of fragments containing the SNPs, SBE primers were designed with 25 bases in core, which bind specifically to the adjacent regions of filtered SNPs, and 5′ non-homologous tails with different length to distinguish SBE products in multiplex reaction. Total DNA extracted from fin clip of 70 fast-growing fish and 70 slow-growing fish was used as template for amplifying fragments containing filtered SNPs. Primer pairs for PCR were re-designed to amplify fragments of equal size. The PCR amplicons of an individual containing filtered SNPs in one multiplex group were pooled, then purified by GeneJET PCR Purification Kit (Thermo Fisher Scientific) according to the manufacturer's instructions. SBE reactions were conducted by ABI SNaPshot Multiplex PCR Kit (Applied BioSystems) according to the manufacture's instruction. The SBE products were purified by enzyme Shrimp Alkaline Phosphatase (SAP) (Thermo Fisher Scientific), then sequenced by ABI3500 Applied BioSystems. The SNP data was analyzed by GeneMapper 5.1 software (Applied BioSystems).

2.2.8. Data analysis

All SNP data of totally 80 individuals of fast-growing fish and 80 individuals of slow-growing fish was analyzed to evaluate the genetic diversity, the nonrandom association between filtered SNPs located in one gene, and to identify candidate SNP markers as well as evaluate the effect of SNPs combination on the growth trait of striped catfish. The genetic diversity was analyzed through polymorphism information content (PIC)

and minor allele frequency (MAF)) by Gene-Calc software. The nonrandom association between SNPs was determined through the linkage disequilibrium – LD calculated by SHEsis software. The SNPs were confirmed as potential markers for growth trait of striped catfish when the corresponding genotypic composition and/or allelic composition between fast- and slow-growing groups were significantly differed. The co-effect of SNPs on the growth trait was analyzed by the significant difference of haplotype's frequencies between fast- and slow-growing groups, which generated from SNPs located in one gene or from SNP markers. Moreover, this co-effect of SNP markers was also evaluated by the significant difference of combined genotype's frequencies between two groups. The significant difference was assessed by p-value < 0.05 from Fisher's exact test, which calculated by SHEsis software and NCSS 2021 Data Analysis (NCSS Statistic Software).

CHAPTER 3. RESULTS

3.1. Structure of genes and proteins of IGF system in striped catfish

The sequence of genes of IGF system in striped catfish had the highest similarity with the sequence of corresponding genes in channel catfish *Ictalurus punctatus*.

IGF1 gene spans 19.4 kb with 4 exons being coding regions for IGF1 protein. This protein contains a peptide signal consisting 33 amino acids, IGF1 domain with 6 conserved cysteine residues and motifs binding to IGFBP, IGF1/2R, insulin receptor type 1, 2. Therefore, the fragments investigated to discover SNP in this gene included exons 1, 2 coding to the peptide signal as well as IGF1 domain, their nearby intronic regions, apart of exon 4 containing the stop codon, 3'-UTR and its previous intron. The *IGF2* gene with 4.9 kb in length includes 4 exons coding to a peptide signal, IGF1 domain and IGF2-C domain from N- to C-terminal. Due to the small size of *IGF2* gene, the whole sequence of 4 exons and their nearby intronic regions was amplified for SNP discovery. With 93.9 kb in size, *IGF1R* gene contains 21 exons, with coding sequence spanning from exon 1 to exon 21. The function domains of IGF1R protein from N- to C-terminal were receptor L domain 1, cysteine rich furine-like domain, receptor L domain 2, Fibronectin type 3 domain and Tyrosine kinase

domain. Because receptor L domain 1 binds to IGF ligand and Tyrosine kinase domain activates the signaling pathway after the binding of IGF1 to IGF1R, the sequences in gene coding to these important functional domains were chosen to amplify for SNP discovery. These amplicons were exon 2, exon 16, 17, 18, 19, 20, 21 and their adjacent intronic regions. The members of IGFBP family, *IGFBP-1 -2, -3 -5 -6* genes, also include 4 exons and 3 introns, with the start codon located in exon 1 and stop codon located in exon 4. The lengths of *IGFBP-1, -2 -3 -5 -6* genes are 2.7 kb, 28.1 kb, 18.4 kb, 12.2 kb and 4.2 kb, respectively. The N-terminal of each protein contains a peptide signal and IGFBP domain, which are encoded by the sequence within exon 1 of corresponding gene. The C-terminal of these proteins also contains Thyroglobulin type 1 domain encoded by the sequence of exon 3 and 4 of their corresponding genes. To discover SNP in *IGFBPs*, fragments coding to amino acid sequence, 5'-UTR, 3'-UTR and their nearby intronic regions of these genes were investigated. The *IGFBP-7* gene is 9 kb in length, including 5 exons with coding sequence spanning from exon 1 to exon 5. IGFBP-7 protein has the similar structure in N-terminal to other IGFBP proteins with a peptide signal and IGFBP domain, which are encoded by sequence within exon 1 of gene. However, IGFBP-7 protein contains KAZAL-FS, Ig3 and Ig domains, not Thyroglobulin type 1 domain in C-terminal. Therefore, the whole sequence of 5 exons and their adjacent intronic regions in *IGFBP-7* gene were amplified to discover SNP.

3.2. The authenticity of Sanger sequence with reference sequence

The similarity between Sanger sequence of fragments in an individual and sequence of reference genes was from 99.9 đến 100 %. This result confirmed the authenticity of Sanger sequence with reference gene in genome sequenced by NGS, contributing the mentioned workflow to discover SNP in candidate genes.

3.3. SNP detecting and filtering in the discovery sample set

The corresponding fragments in 10 fast-growing fish and 10 slow-growing fish were sequenced by Sanger sequencing and aligned to the reference sequence to detect SNPs. The detected SNPs were then filtered according to criteria mentioned in 2.2.6.

3.3.1. Detecting and filtering SNP in IGF1 gene

A total of 10 SNPs were detected in introns of *IGF1* gene (Table 3.1). In which, 4 SNP 13185 C>T, 13263 T>C located in intron 1 and 13680 A>T, 13684 G>C located in intron 2 had the SNP ratio in one group be at least 0.3 (Table 3.1). However, only SNP 13680 A>T showed the significant difference in genotypic and allelic composition between two groups, with p value was 0.03 and 0.005, respectively (Table 3.1), thus this SNP was filtered to validate in the validation sample set.

Table 3.1. Discovered SNPs in *IGF1* gene

Nº	**Position of SNP in gene**	**Ref**	**Alt**	**Genetic composition**		**Allelic composition**	
				FG (ª)	**SG (ª)**	**FG**	**SG**
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

**: minor allele, FG: fast-growing group, SG: slow-growing group, (ª): the SNP ratio in each group, (2), (3), (4): filtered SNP passed the criteria number 2, 3, 4.*

3.3.2. Detecting and filtering SNP in IGF2 gene

There were 12 SNPs detected in *IGF2* gene. Although 2 SNPs 1355 A>G, 1391 G>A located in intron 2 and SNP 2061 T>A in intron 3 had the SNP ratio in one group be at least 0.3, these SNPs did not present the significant difference in genetic composition as well as allelic composition between fast- and slow-growing groups. Therefore, no SNP in *IGF2* gene was filtered to further validate in the next procedure.

3.3.3. Detecting and filtering SNP in IGF1R gene

All of 9 SNPs detected in *IGF1R* gene were located in introns, and had the SNP ratio in one group be at least 0.3 (Table 3.3). In which, 13357 T>C, 15392 T>A and 83894 A>G showed the significant difference in genotypic composition and/or allelic composition between fast- and slow-growing groups (Table 3.3), hence, these SNPs were filtered to validate in the validation sample set.

Table 3.3. Discovered SNPs in *IGF1R* gene

Nº	**Position of SNP in gene**	**Ref**	**Alt**	**Genetic composition**		**Allelic composition**	
				FG ([a])	**SG ([a])**	**FG**	**SG**
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	

[*]: minor allele, FG: fast-growing group, SG: slow-growing group, ([a]): the SNP ratio in each group, [(2), (3), (4)]: filtered SNP passed the criteria number 2, 3, 4.

3.3.4. Detecting and filtering SNP in IGFBP-1 gene

There were 4 SNPs detected in *IGFBP-1* gene, including 1 SNP located in promoter and 3 SNPs located in introns. However, none of these SNPs passed the criteria to be filtered.

3.3.5. Detecting and filtering SNP in IGFBP-2 gene

A total of 10 SNPs were detected in *IGFBP-2* gene, including 1

synonymous SNP located in coding sequence and 9 SNPs located in non-coding region. Nevertheless, none of these SNPs passed the filtering criteria.

3.3.6. Detecting and filtering SNP in IGFBP-3 gene

There were 10 SNPs detected in *IGFBP-3* gene, including 6 SNPs located in introns and 4 SNPs located in 5'UTR, coding sequence and 3'-UTR. SNP 704C>G located in exon 1 (Figure 3.9) was a non-synonymous variant, causing the substitution of amino acid in IGFBP-3 protein sequence (p.Leu8Val). Therefore, this SNP was filtered to validate in the validation sample set. Other SNPs located in non-coding sequence did not pass the criteria to be filtered.

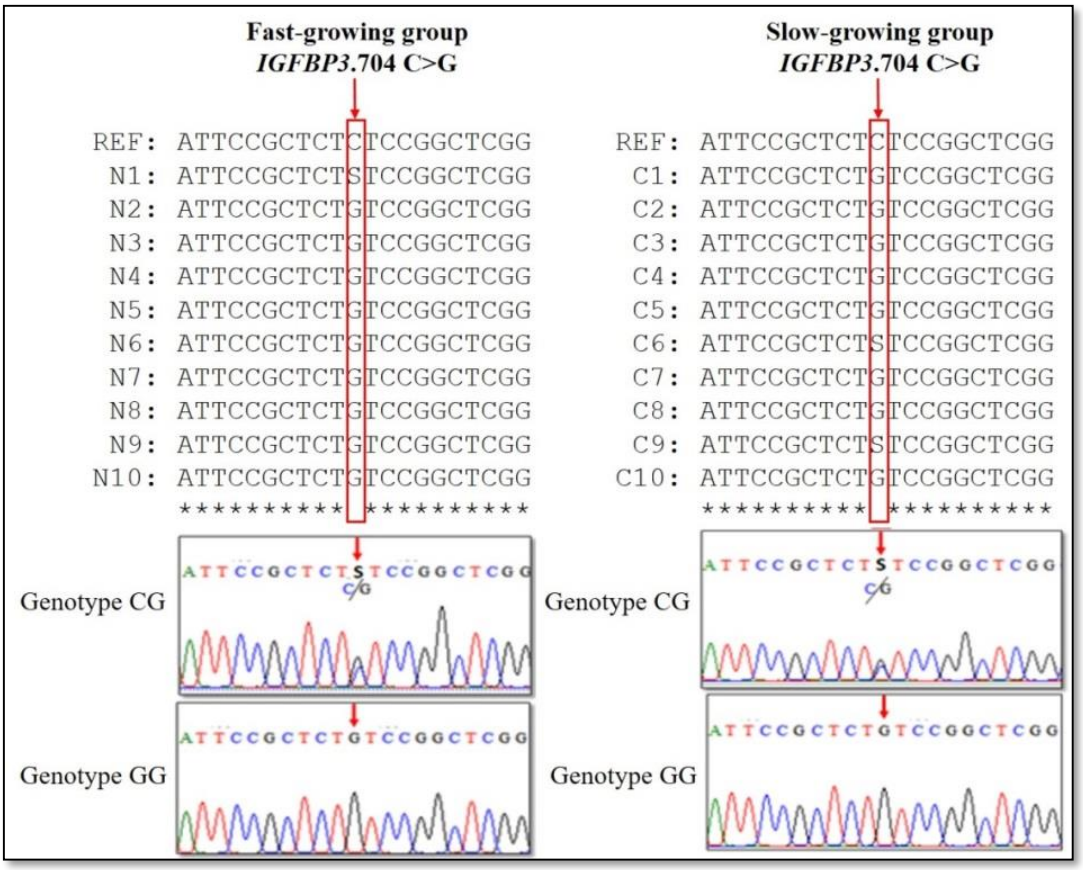

Figure 3.9: Detecting SNP *IGFBP-3*.704 C>G in *IGFBP-3* gene by aligning Sanger sequence of corresponding fragments with reference sequence.

REF: reference sequence. N1 - N10: Sanger sequences of 10 fast-growing fish. C1 - C10: Sanger sequences of 10 slow-growing fish. Red arrow and red border: position of detected SNP in each individual.

3.3.7. Detecting and filtering SNP in IGFBP-5 gene

All SNPs detected in *IGFBP-5* gene were located in exons. In which, SNP 525T>A in exon 1 caused the substitution of amino acid in IGFBP-5 protein sequence (Val16Glu). Therefore, this non-synonymous SNP was filtered to validate in the validation sample set (Figure 3.10). Other SNPs did not pass the criteria to be filtered.

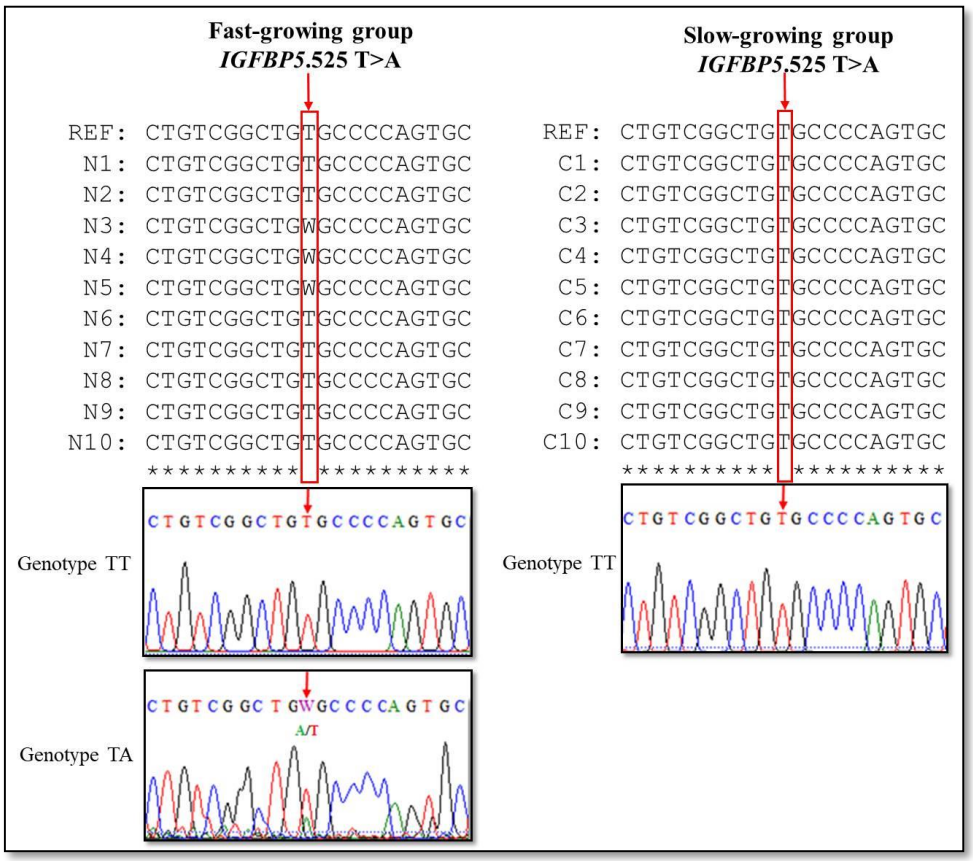

Figure 3.10. Detecting SNP *IGFBP-5*.525 T>A in *IGFBP-5* gene by aligning Sanger sequence of corresponding fragments with reference sequence.

REF: reference sequence. N1 - N10: Sanger sequences of 10 fast-growing fish. C1 - C10: Sanger sequences of 10 slow-growing fish. Red arrow and red border: position of detected SNP in each individual.

3.3.8. Detecting and filtering SNP in IGFBP-6 gene

A total of 17 SNPs detected in *IGFBP-6* gene were located in non-coding region. In which, 7 SNPs had the SNP ratio in one group be at least 0.3, including SNPs 560 A>G, 2275 C>A, 2278 C>A, 2423 G>T, 3177 T>C, 3469 T>C and 3505 G>A (Table 3.8, quote). However, only SNP 2278 C>A was filtered because of the significant difference in genotypic composition of this SNP between two groups (p=0.02) (Table 3.8, quote).

Table 3.8. Discovered SNPs in *IGFBP-6* gene (quote)

Nº	Position of SNP in gene	Ref	Alt	Genetic composition		Allelic composition	
				FG (ª)	SG (ª)	FG	SG
1	I1_560(2)	A	G*	6AA:2AG:2GG (0.50)	8AA:1AG:1GG (0.22)	14A:6G	17A:3G
				p = 0.52		p = 0.45	
11	I3_2275(2)	C	A*	8CC:2AA (0.25)	5CC:5AA (1.00)	16C:4A	10C:10A
				p = 0.35		p = 0.09	
12	**I3_2278(2,4)**	**C**	**A***	**8CC:2AA (0.25)**	**4CC:6AA (1.50)**	**16C:4A**	**8C:12A**
				p = 0.17		**p = 0.02**	
13	I3_2423(2)	G	T*	6GG:4TT (0.67)	4GG:6TT (1.50)	12G:8T	8G:12T
				p = 0.66		p = 0.34	
15	E4_3'UTR 3177(2)	T	C*	6TT:4CC (0.67)	6TT:4CC (0.67)	12T:8C	12T:8C
				p = 1.0		p = 1.0	
16	E4_3'UTR 3469(2)	T*	C	3TT:2TC:5CC (0.71)	1TT:9CC (0.11)	8T:12C	2T:18C
				p = 0.13		p = 0.06	
17	E4_3'UTR 3505(2)	G	A*	7GG:3AA (0.43)	6GG:4AA (0.67)	14G:6A	12G:8A
				p = 1.0		p = 0.74	

**: minor allele, FG: fast-growing group, SG: slow-growing group, (ª): the SNP ratio in each group, (2), (4): filtered SNP passed the criteria number 2, 4.*

3.3.9. Detecting and filtering SNP in IGFBP-7 gene

A total of 27 SNPs were detected in *IGFBP-7* gene, including 2 non-synonymous SNP 344 T>C (p.Leu78Pro) and 4559 C>A (p.Leu189Met) (Table 3.9, quote).

Table 3.9. Discovered SNPs in *IGFBP-6* gene (quote)

Nº	Position of SNP in gene	Ref	Alt	Genetic composition		Allelic composition	
				FG (ª)	SG (ª)	FG	SG
2	**E1 - CDS 344 (1)**	**T***	**C**	1TT:2TC:7CC	3TT:3TC:4CC	4T:16C	9T:11C
12	**I2 - 2060(2,3,4)**	**A**	**G***	**10GG (0.00)**	**3AA:2AG:5GG (1.40)**	**20G**	**8A:12G**
				p = 0.02		p = 0.003	
20	**E3 - CDS 4559 (1)**	**C***	**A**	**2CC:3CA:5AA**	**10AA**	**7C:13A**	**20A**

**: minor allele, FG: fast-growing group, SG: slow-growing group, (ª): the SNP ratio in each group, (1), (2), (3), (4): filtered SNP passed the criteria number 1, 2, 3, 4.*

Among other SNPs, 17 SNPs had the SNP ratio in one group be at least 0.3, including SNP -376 T>G in promoter, SNPs 561 G>A, 598 C>A, 1052 T>A, 1117 C>G, 1160 G>C and 1216 C>A in intron 1, SNPs 2060 A>G, 2138 T>C, 3981 T>G, 4058 A>G, 4334 G>A and 4465 C>T in intron 2, SNP 4685 T>C in intron 3, 7405 C>T, SNPs 7676 G>C and 7692 T>C in intron 4. However, only SNP 2060 A>G presented the significant difference in genotypic composition/allelic composition between fast- and slow- growing group (p-value < 0.05) (Table 3.9, quote). Thus, 3 SNPs 344 T>C (p.Leu78Pro), 2060 A>G and 4559 C>A (p.Leu189Met) were filtered to further validate in validation sample set.

3.4. Association analysis of SNPs with growth traits

3.4.1. Individual SNP genotyping in validation sample set

Fragments containing 10 filtered SNPs in *IGF1, IGF1R, IGFBP-3,-5 -6, -7* genes of 140 individuals in validation sample set were amplified and purified to be used as template for multiplex SBE reactions genotyping SNP. SNP data collected from 160 individuals of both discovery and validation sample sets was used to analyze the association of SNPs with growth traits.

3.4.2. Analyzing the association between filtered SNPs and growth traits of striped catfish

3.4.2.1. Filtered SNPs in IGF1 gene

Table 3.11. Analyzing the association between SNP 13680 A>T in *IGF1* gene with growth traits of striped catfish

Parameters	**Fast-growing group**	**Slow-growing group**	**p**
Genotypic composition ([b])	AA (18) AT (61) TT (01)	AA (07) AT (64) TT (07) NN (02)	0.009**
Allelic composition ([c])	A (97) T (63)	A (78) T (78)	0.06
PIC	0.494		
MAF	0.446		

([b]): The number of individuals carrying genotype in each group, *NN:* Non-identified genotype individuals, *([c]):* The number of alleles in each group, ***: p < 0.01.*

The statistical difference in genetic composition generated from SNP *IGF1*.13680 A>T between fast- and slow- growing groups (p= 0.009) (Table 3.11) suggested the significant association of this SNP with growth traits.

3.4.2.2. Filtered SNPs in IGF1R gene

Among 3 filtered SNPs in *IGF1R* gene 13357 T>C, 15392T>A and 83894 A>G, only SNP 13357 T>C showed the significant difference in genetic composition between fast- and slow- growing group (p= 0.03) (Table 3.12).

Table 3.12. Analyzing the association between 3 filtered SNPs in *IGF1R* gene with growth traits of striped catfish.

SNP name	Genetic compositon ([b])			Allelic composition ([c])			PIC	MAF
	FG	SG	p	FG	SG	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

*FG: Fast-growing group, SG: Slow-growing group, ([b]): The number of individuals carrying genotype in each group, NN: Non-identified genotype individuals, ([c]): The number of alleles in each group, *: $p < 0.05$.*

Table 3.14. Analyzing the association between haplotypes generated from 3 filtered SNPs in *IGF1R* gene and growth traits of striped catfish.

Haplotype generated from 3 SNPs				FG ([d])	SG ([d])	P
13357 T>C	15392 T>A	83894 A>G	ID			
C	**A**	**A**	***IGF1R*_H1**	**35.39 (0.239)**	10.18 (0.068)	**2.73e-5****
C	**T**	**G**	*IGF1R*_H2	3.77 (0.025)	**46.79 (0.312)**	**7.47e-11****
T	**A**	**A**	*IGF1R*_H3	28.96 (0.196)	**54.78 (0.365)**	**0.001****
T	**T**	**A**	***IGF1R*_H4**	**11.15 (0.075)**	1.01(0.007)	**0.002****
T	**T**	**G**	***IGF1R*_H5**	**62.59 (0.423)**	34.17 (0.228)	**0.00018****

*FG: Fast-growing group, SG: Slow-growing group, ([d]): The frequency of haplotype in each group, which haplotype having the frequency smaller than 0.03 in both two groups was out of analysis, **$p<0.01$, ***$p<0.001$.*

Haplotype analysis showed 3 haplotypes *IGF1R*_H1 (CAA), H4 (TTA) and H5 (TTG) were significant associated to fast-growing group ($p<0.01$), while 2 haplotypes H2 (CTG) and H3 (TAA) were significant associated to slow-growing group ($p<0.01$) (Table 3.14).

3.4.2.3. Filtered SNPs in IGFBP-3 gene

The genotypic composition generated from SNP 704 C>G (Leu8Val) between fast- and slow-growing groups significantly differed ($p<0.05$) (Table 3.15). Moreover, there was a significant difference in the allelic

composition generated from this SNP between two groups (p<0.05) (Table 3.15).

Table 3.15. Analyzing the association between SNP 704 C>G (p.Leu8Val) in *IGFBP-3* gene with growth traits of striped catfish

Parameters	**Fast-growing group**	**Slow-growing group**	**P**
Genotypic composition ([b])	CG (45) GG (35)	CG (26) GG (53) NN (01)	0.03*
Allelic composition ([c])	C (45) G (115)	C (26) G (132)	0.013*
PIC	0.347		
MAF	0.223		

*([b]): The number of individuals carrying genotype in each group, NN: Non-identified genotype individuals, ([c]): The number of alleles in each group, *: p < 0.05.*

3.4.2.4. Filtered SNPs in IGFBP-5 gene

The genotypic composition and allelic composition generated from SNP 525 T>A (p.Val16Glu) significantly differed between fast- and slow-growing groups (p < 0.001) (Table 3.16).

Table 3.16. Analyzing the association of SNP 525 T>A (p.Val16Glu) in *IGFBP-5* gene with growth traits of striped catfish

Parameters	**Fast-growing group**	**Slow-growing group**	**P**
Genotypic composition ([b])	TT (07) TA (69) AA (03) NN (01)	TT (41) TA (37) NN (02)	1.13e-8***
Allelic composition ([c])	T (83) A (75)	T (119) A (37)	1.14e-5***
PIC	0.459		
MAF	0.357		

*([b]): The number of individuals carrying genotype in each group, NN: Non-identified genotype individuals, ([c]): The number of alleles in each group, **: p < 0.01.*

3.4.2.5. Filtered SNPs in IGFBP-6 gene

The validation of filtered SNP 2278 C>A in *IGFBP-6* gene showed any statistical difference in genotypic composition/allelic composition between fast- and slow- growing groups.

3.4.2.5. Filtered SNPs in IGFBP-7

SNP 344 T>C (p.Leu78Pro) did not present the significant difference in genotypic composition/ allelic composition between fast- and slow-growing groups. There was significant difference in genotypic

composition/ allelic composition generated from SNP 2060 A>G between two groups (all p-value < 0.001). The genotypic composition generated from SNP 4559 C>A (p.Leu189Met) between two groups significantly differed (p < 0.05) (Table 3.18). Therefore, 2 SNP 2060 A>G and 4559 C>A (p.Leu189Met) were two potential candidate SNP markers for growth traits of striped catfish.

Table 3.18. Analyzing the association between 3 filtered SNPs in *IGFBP-7* gene with growth traits of striped catfish.

SNP name	**Genetic compositon ([b])**			**Allelic composition ([c])**			**PIC**	**MAF**
	FG	**SG**	**p**	**FG**	**SG**	**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

FG: Fast-growing group, SG: Slow-growing group, ([b]): The number of individuals carrying genotype in each group, NN: Non-identified genotype individuals, ([c]): The number of alleles in each group, *: p < 0.05 **: p < 0.01, ***: p < 0.001.

Table 3.20. Analyzing the association between haplotypes generated from 3 filtered SNPs in *IGFBP-7* gene and growth traits of striped catfish.

Haplotype generated from 3 SNPs				**FG ([d])**	**SG ([d])**	**P**
344 T>C	**2060 A>G**	**4559 C>A**	**ID**			
C	**A**	**A**	*IGFBP7*_H1	20.32 (0.127)	**58.84 (0.377)**	**2.97e-7*****
C	**A**	**C**	*IGFBP7*_H2	2.75 (0.017)	**22.13 (0.142)**	**3.97e-5*****
C	**G**	**A**	***IGFBP7*_H3**	**66.69 (0.417)**	23.17 (0.148)	**1.31e-7*****
C	**G**	**C**	***IGFBP7*_H4**	**12.25 (0.077)**	1.86 (0.012)	**0.0054****
T	A	A	*IGFBP7*_H5	30.93 (0.193)	26.02 (0.167)	0.54
T	G	A	*IGFBP7*_H6	27.07 (0.169)	23.97 (0.154)	0.708

FG: Fast-growing group, SG: Slow-growing group, ([d]): The frequency of haplotype in each group, which haplotype having the frequency smaller than 0.03 in both two groups was out of analysis, **p<0.01, ***p<0.001.

Generated from 3 filtered SNPs in *IGFBP-7* gene, haplotypes *IGFBP-7*_H1 (CAA) and H2 (CAC) were significantly associated to slow-growing group (p < 0.001), while haplotypes *IGFBP-7*_H3 (CGA) and H4

(CGC) were significantly associated to fast- growing group ($p < 0.01$) (Table 3.20).

3.4.2.6. The cumulative effect of SNP markers on the growth traits of striped catfish

The cumulative effect of 6 SNP markers including *IGF1*.13680 A>T, *IGF1R*.13357 T>C, *IGFBP-3*.704C>G, *IGFBP-5*.525T>A, *IGFBP-7*.2060 A>G and *IGFBP-7*.4559 C>A on the growth traits was analyzed by the significant difference in haplotype frequency and combined genotype frequency between fast- and slow- growing group.

Table 3.21. Analyzing the association of haplotypes generated from 6 SNP markers with growth traits of striped catfish

Haplotype generated from 6 SNP markers						**ID**	**FG ([d])**	**SG ([d])**	**p**
***IGF1* 13680 A>T**	***IGF1R* 13357 T>C**	***IGFBP3* 704 C>G**	***IGFBP5* 525 T>A**	***IGFBP7* 2060 A>G**	***IGFBP7* 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, FG: Fast-growing group, SG: Slow-growing group, ([d]): The frequency of haplotype in each group, which haplotype having the frequency smaller than 0.03 in both two groups was out of analysis, *p<0.05, **p<0.01, ***p<0.001*

Haplotype analysis showed 4 haplotypes CO_H2, H5, H11, H14 prevailed in fast- growing group, while the frequencies of 7 haplotypes CO_H1, H3, H7, H8, H9, H10, H14 were higher in slow-growing group (Table 3.21).

Generated from 6 SNP markers, combined genotype AT, CT, GC, AT, GA and AA (also performed by haplotype CO_H2 and CO_H11 (diplotype H2/H11)) had the higher frequencies in fast- growing group, and combined genotypes AT, TT, GC, AT, GG and AA (also performed by haplotype CO_H2 and CO_H5 (diplotype H2/H5) only presented in fast-growing group ($p<0.05$) (Table 3.22).

Table 3.22. Analyzing the association of combined genotype generated from 6 SNP markers with the growth traits of striped catfish

Combined genotype generated from 6 SNP markers						**FG (^{e})**	**SG (^{e})**	**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**			
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}): Frequency of combined genotype in each group, which combined genotype having the frequency smaller than 3% in both two groups was out of analysis, $^*p<0,05$, $^{**}p<0,01$.

CHAPTER 4. DISCUSSIONS

4.1. The similarity between sequence as well as structure of genes in IGF system of striped catfish and that of teleost fish

The candidate genes in IGF system of striped catfish had the highest similarity in sequence with corresponding genes of channel catfish *Ictalurus*

punctatus, which are very closed to striped catfish in the phylogeny and also classified into the order *Siluriformes* catfish. The structure of IGF1, IGF1R, IGFBP-1, -2 -3,-5,-6,-7 genes/proteins of striped catfish was similar to the structure of corresponding genes/proteins of other teleost fish, and contained conserved functional domains.

4.2. The genetic diversity of genes in IGF system of striped catfish

Non-coding SNPs accounted for 93.2% of total detected SNPs in this study. This result was consistent with the findings of previous researches in genome showing that the proportion of non-coding SNP was approximate 90%. The similar frequency of non-coding SNP was also indicated in studies of candidate genes related growth traits, such as Growth Hormone (*GH*) gene in mandarin fish *Siniperca chuatsi*, *IGF1* gene in common carp *C. carpio*, *myostatin-1* gene in *Ancherythroculter nigrocauda*, *GH*, *IGF1* and *Myogenin* (*MyoG*) genes in Nile tilapia *O. niloticus*. In constrast, synonymous SNPs in this study took only 3.8%, this low proportion could be explained by the functional constraint of amimo acid changes, which should be eliminated by the evolution. Consisting to this phenomenon, some researches in teleost fish presented the low percentage of non-synonymous SNP in total of discovered SNPs, at 4% in Atlantic salmon *S. salar*, 10% in chanel catfish *I. punctatus*, and in sole *Solea solea L.*, 11% to 14% in turbot *Scophthalmus maximus*.

Only 9.7% detected SNPs (10/103 SNPs) was filtered to individual genotyping in validation sample set. This filtration was necessary to ensure accuracy while still spending time and research cost. Many researches also filtered the large number of detected SNP to choose candidate SNPs, which further validate the association of SNPs and the traits of interest in the workflow. For examples, a panel of 60 SNPs sourced from a pool of 7000 SNPs in Atlantic salmon *S. salar* was used to diagnose salmon as being farmed or wild. In blue catfish *Ictalurus furcatus*, 64 in total of 4275 SNPs were chosen to individual genotyping to determine wild and domesticated populations. Recently, the growth traits of mandarin fish were evaluated by 48 SNPs in total of 5205 SNPs found by genome –wide association study.

None of SNPs was filtered in *IGF2*, *IGFBP-1* and *IGFBP-2* genes of striped catfish. In the literature, although the expression level of these

genes was indicated associating to the growth of channel catfish *I. punctatus*, half-smooth tongue sole, *Cynoglossus semilaevis*, Nile tilapia *O. niloticus*, European sea bass *D. labrax*, species of Cyprinus, and zebra fish *Danio rerio*, there was a few or none of studies showing the significant association of SNPs in these genes with the growth traits of teleost fish.

In this study, 4 synonymous SNPs were discovered in candidate genes, including SNP *IGFBP-3*. 704C>G (Leu8Val), *IGFBP-5*. 525T>A (Val16Glu), *IGFBP-7*. 344 T>C (Leu78Pro) and *IGFBP-7*. 4559 C>A (Leu189Met). In which, the substituted amino acids Leu8Val in IGFBP-3 and Val16Glu in IGFBP-5 were *in silico* predicted locating in the peptide signal at N terminus of corresponding protein. Because the signal peptide is crucial for the transport and secretion of IGFBP-3 and IGFBP-5, the effect of these two non-synonymous SNPs on the function of corresponding proteins needs to be further clarified. The alignment of IGFBP-7 protein sequences in striped catfish and other teleost fish showed the Leu78Pro caused by SNP 344T>C was not conserved residue, while Leu189Met caused by SNP 4559C>A was located in conserved region of immune domains Ig and Ig3.

4.3. The association between the genetic variations of genes in IGF system and the growth traits of striped catfish

SNP 13680 A>T in intron 2 of *IGF1* gene showed the significant association with the growth traits of striped catfish. This characteristic contributed the association of non-coding SNPs in *IGF1* gene and the growth, similar to results of previous researches in teleost fish such as common carp *C. carpio*, Atlantic salmon *S. salar*, and in other taxa including cattles and seedcracker. In *IGF1R* gene of striped catfish, only SNP 13357 T>C in total of 3 filtered SNPs still exhibited the significant difference between the fast- and slow-growing groups. Research on the genetic variation of *IGF1R* gene in endemic fish *O. potamophila* in China also identified 3 SNPs in non-coding region, and found one SNP (1208G > A) was significantly associated with important growth traits. In literature, almost SNPs in *IGFBP-3* and *IGFBP-5* genes relating to the growth traits were located in non-coding region, such as promoter in human, swine, common carp, chicken, and introns in swine. As a different highlight, this

study showed 2 non-synonymous SNPs *IGFBP-3*.704C>G (p.Leu8Val) and *IGFBP-5*.525T>A (p.Val16Glu) in coding sequence of genes being significantly associated to the growth of striped catfish. This result was consolidated by the observation that non-synonymous SNPs screened in other candidate genes were proven relating to the growth of aquatic species such as razor clam, hybrid of *Culter alburnus* (♀) x *Ancherythroculter nigrocauda* (♂) individuals, and channel catfish. In this study, there was only one SNP 2278C>A in total of 17 detected SNPs in *IGFBP-6* gene being filtered to validate, however, this SNP was not associated to the growth of striped catfish. This result fitted the observation that *IGFBP-6* genes of teleost fish were rather understudied, with unclear roles and functions. In *IGFBP-7* gene, intronic SNP 2060 A>G and non-synonymous SNP 4559 C>A (p.Leu189Met) were significantly associated to the growth traits of striped catfish, contributing the significant association of intronic SNPs as well as non-synonymous SNPs with the growth traits of teleost fish as previous studies suggested. The genetic diversity of 6 SNP markers evaluated by PIC value indicated the moderate genetic diversity of these SNPs (PIC from 0.25 to 0.5), suggesting the reasonable potential for breeding selection. Be determined as common variant with MAF value larger than 20%, almost of 6 SNP markers significantly contributed to the genetic variance. On the other hand, SNP *IGFBP-7*. 4559C>A (p.Leu189Met) showed lower PIC and MAF values. This phenomenon could be explained that this SNP caused the substitution of amino acid in the conserved region of immune domains Ig and Ig3 of corresponding protein.

In *IGF1R* gene, although two filtered SNPs 15392T>A and 83894A>T were not significantly associated with the growth trait of striped catfish, haplotype analysis of the three filtered SNPs in this gene revealed that the haplotype combination was significantly associated to the growth. The similar results were indicated in *IGFBP-7* gene when the haplotypes generated from non-associated SNP 344T>C (p.Leu78Pro) and two SNP markers 2060A>G, 4559C>A (p.Leu189Met) were significantly associated to the growth traits of striped catfish. This observation indicated the limit genetic information of individual SNPs but the more meaningful association of haplotype combinations to the growth traits, being

consistent with previous studies in largemouth black bass *Micropterus salmoides*, channel catfish *I. punctatus* and Atlantic salmon *S. salar*.

4.4. The development of SNP markers for growth selection of striped catfish

The cumulative effect of 6 SNP markers on the growth traits of striped catfish was analyzed by the significant difference in frequency of haplotype as well as combined genotype between fast- and slow- growing group. The consistency of results in this study provided potential haplotype/combined genotype markers associated to the growth traits of striped catfish. Moreover, the growth traits of striped catfish should be affected by many difference genes, hence molecular markers for growth traits of striped catfish should be developed based on the combinations of potential SNPs as previously reported in rainbow-trout *Oncorhynchus mykiss* and common carp *C. carpio*.

CONCLUSIONS

1. The sequences and structures of 9 genes in IGF system including *IGF1*, *IGF2*, *IGF1R*, *IGFBP-1*, *-2*, *-3*, *-5*, *-6*, *-7* genes were analyzed. The important fragments in the candidate genes were determined and sequenced.
2. Total of 103 SNPs were detected in 9 genes (*IGF1, IGF2, IGF1R, IGFBP-1, -2, -3, -5, -6, -7)* of striped catfish individuals of discovery sample set, and 10 SNPs were filtered to validate in the bigger population.
3. The filtered SNPs were individual genotyping in validated population by Single base extension method. Six SNP markers significantly associated to the growth traits of striped catfish, including *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 and *IGFBP-7*. 4559 C>A (p.Leu189Met).
4. Total of 11 haplotypes and 2 combined genotypes formed by 6 SNP markers significantly associated to the growth traits of striped catfish.

REQUESTS

The association between growth traits of striped catfish and the SNP markers as well as generated haplotypes, diplotypes, combined genotypes needs to be further validated in the bigger population, aims to use these genetic variations as marker-assisted selection for growth traits of striped catfish.

THE NEW FINDING OF THE DOCTORAL THESIS

1-Identified 6 candidate SNP markers in 9 genes in IGF system for growth traits of striped catfish.

2-Identified haplotypes, combined genotypes generated by SNPs associated significantly with growth traits of striped catfish.

LIST OF PUBLICATIONS

1. Le Thi Nguyen Binh, Nguyen Thi Hoa, **Tran Thi Huyen Trang**, Nguyen Thanh Phuong, Kim Thi Phuong Oanh. *Structure of insulin-like growth factor 2 (IGF2) gene from striped catfish (Pangasianodon hypophthalmus)***.** Vietnam Journal of Biotechnology, 2019, 17(3): 455-463.
2. **Tran Thi Huyen Trang,** Le Thi Nguyen Binh, Nguyen Thi Hoa, Tran Son Hoang, Kim Thi Phuong Oanh. *Structure of Insulin-like Growth Factor Binding Protein (IGFBP) gene from striped catfish (Pangasianodon hypophthalmus)*. Proceeding National Biotechnology Conference 2019, Ho Chi Minh City, Vietnam, 2019:17-22.
3. **Trang Thi Huyen Tran**, Hoa Thi Nguyen, Binh Thi Nguyen Le, Phuc Huu Tran, Sang Van Nguyen, Oanh Thi Phuong Kim. *Characterization of single nucleotide polymorphism in IGF1 and IGF1R genes associated with growth traits in striped catfish (Pangasianodon hypophthalmus Sauvage, 1878).* Aquaculture, 2021, 538: e736542 (doi: 10.1016/j.aquaculture.2021.736542. Epub 2021 May 30)
4. **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim. *Significant association between a non-synonymous SNP in IGFBP5 gene and the growth of striped catfish (Pangasianodon hypophthalmus, Sauvage, 1878)*. Vietnam Journal of Biotechnology, 2023, 21(1): 1-11 (doi: 10.15625/1811-4989/17647)
5. **Trang Thi Huyen Tran**, Hoang Son Tran, Binh Thi Nguyen Le, Sang Van Nguyen, Hai-Anh Vu, Oanh Thi Phuong Kim. *Novel single nucleotide polymorphisms of Insulin-like Growth Factor Binding Protein 7 (IGFBP7) gene significantly associated with growth traits in striped catfish (Pangasianodon hypophthalmus Sauvage, 1878)*. Molecular Genetics and Genomics, 2023, 298(4): 883-893 (doi:10.1007/s00438-023-02016-2. Epub 2023 Apr 25)
6. **Trang Thi Huyen Tran**, Binh Thi Nguyen Le, Sang Van Nguyen, Oanh Thi Phuong Kim**.** Non-synonymous polymorphism in *igfbp-3* gene associated with growth traits in striped catfish (*Pangasianodon hypophthalmus*, Sauvage, 1878). Vietnam Journal of Biotechnology, accepted.