Genetic analysis of loci controlling heat tolerance and ESC disease resistance in catfish

by

Yulin Jin

A dissertation submitted to the Graduate Faculty of
Auburn University
in partial fulfillment of the
requirements for the Degree of
Doctor of Philosophy

Auburn, Alabama
Aug 4, 2018

Keywords: Heat stress; ESC; Disease resistance; GWAS; BSR-Seq; Catfish

Approved by

Zhanjiang Liu, Chair, Professor, School of Fisheries, Aquaculture and Aquatic Sciences
Rex Dunham, Co-chair, Professor, School of Fisheries, Aquaculture and Aquatic Sciences
Nannan Liu, Professor, Department of Entomology and Plant Pathology
Charles Chen, Professor, Department of Crop, Soil and Environmental Sciences

Abstract

Catfish is one of the main aquaculture species in the US. However, the catfish industry is under the threat of various of environmental stresses and diseases, with heat stress and ESC are two major representatives. One promising strategy for reducing the magnitude of the threats is to select and develop catfish lines that are genetically resistant to heat stress and ESC disease. Understanding the fundamental mechanism conferring tolerance to heat stress and ESC disease have been studies for decades, and it is of great essential for genetic enhancement programs. However, it is still far from unrevealing the causal genes and genomics locus responsible for these performance traits. In this dissertation, I aim to characterize genes and genomic loci controlling heat tolerance and ESC disease resistance in catfish using integrated genetic, genomic and transcriptomic analyses.

Firstly, a genome-wide association study (GWAS) identified three significant SNP markers conferring response to heat stress at the genome-wide significance level in F2 hybrid catfish. The SNP located on linkage group 14 explained 12.1% of phenotypical variation. The other two SNPs located on linkage group 16 explained 11.3% and 11.5% of phenotypical variation, respectively. A total of 14 genes with heat stress related functions were detected within the significant associated regions, with the centrality of genes involved in protein degradation process through ubiquitination pathway. Secondly, a bulked segregant RNA-Seq (BSR-Seq) analysis was conducted to profile DEGs and map ESC resistant QTLs in channel catfish using both liver and intestine tissues. Transcriptomic profiling analysis revealed divergent cellular responses between liver and intestine

after infection, demonstrating they could collaborate closely while keep hemostasis using own specific strategies and potential unique mechanisms after infection. Five significant SNPs with large ED values were identified, with three significant SNPs were from liver and physically linked on LG1, the other two were from intestine and located on LG12 and LG26, respectively. Collectively, these significant SNPs suggest three associated QTLs with ESC resistance, which are successfully verified our previous GWAS studies. Eleven genes were found to be differentially regulated between resistant fish and susceptible fish within the three QTL regions, indicating their important involvements in disease resistance. Of particular interest is the Apo-14 kDa gene displayed both differentially up-regulation and significantly allelic segregation of SNP between resistant fish and susceptible fish, indicating Apo-14 kDa could be a promising candidate gene involved in ESC resistance.

Overall, my research identified significantly SNPs, genomic regions and potential candidate genes associated with heat stress and ESC resistance by using GWAS and BSR-Seq, respectively. GWAS and BSR-Seq are approved as two major efficient and powerful approaches to mapping genomic loci responsible for many catfish performance traits. The associated SNPs could be promising candidates for selecting heat-tolerant or ESC resistance catfish lines after validating their effects on larger and various catfish populations. The isolation, functional study and regulation networks of these potential candidate genes will be the focus in future study, which will facilitate the better knowledge of heat stress and ESC resistance in catfish.

Acknowledgments

First and foremost, I would like to express my special appreciation and thanks to my major advisor Professor Dr. Zhanjiang Liu for the continuous support of my Ph.D study and related research, for his patience, motivation, and immense knowledge. I would like to thank you for encouraging my research and for allowing me to grow as a research scientist. I would also like to thank my co-advisor Dr. Rex Dunham provided valuable knowledge and expertise for the fish biology and genetics, allowing me to combine my love for the exciting world of catfish. I also would like to appreciate and thank the rest of my committee members: Prof. Nannan Liu and Prof. Charles Chen, and the dissertation outside reader Dr. Mark Liles, for their insightful comments and encouragement. I would especially like to thank all the members of the Fish Molecular Genetics and Biotechnology Laboratory, past and present, who have helped me with my research. I am especially grateful to Dr. Dongya Gao for supporting me spiritually throughout these 4 years of graduate studies at Auburn University.

A special thanks to my family for their unconditional love and support of my academic goals. My special thanks also goes to Dr. Stewart Schneller and his wife Aina for providing me with both academic and experiences to live and study abroad. I sincerely thank all my dear friends for their friendship, encouragement, and moral support. Lastly, I would like express appreciation to my financial support came from Chinese scholarship Council.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

ESC	Enteric Septicemia of Catfish
GWAS	Genome Wide Association Studies
QTL	Quantitative Trait Locus
LD	Linkage Disequilibrium
BSR-Seq	Bulked Segregant RNA-Seq
BSA	Bulk Segregant Analysis
ED	Euclidean Distance
DEG	Differentially Expressed Gene

Chapter 1 Literature review

Heat stress and the effects on catfish industry

Global warming and the effects on fisheries

Greenhouse gas emissions have altered mean annual temperatures, precipitation and weather patterns. The Intergovernmental Panel on Climate Change (IPCC 2001) employed several global circulation models (GCMs) and predicted an increased likelihood of 1-7℃ in mean global temperature over the next hundred years [1]. Ongoing global climate change is predicted to affect organisms during all life stages, thereby affecting populations of a species, communities and the functioning of ecosystems [2]. In the freshwater systems, the general effects of climate change on environmental variables will likely be increased water temperature, decreased dissolved oxygen and the increased toxicity of pollutants. Therefore, as it strengthens over time, global climate change will become a more powerful stressor for fish living in natural or artificial systems. Temperature is a major seasonal environmental factor that can undergo daily fluctuations and short erratic lows and highs. Each species of the aquatic ectotherms has evolved physiologically to live within a specific range of environmental variation, and existence outside of that range can be stressful or fatal [3]. The acclimation temperature (both constant and cyclic), magnitude and direction of the temperature shift, frequency of temperature change and rate of temperature change can have important effects on their life history [4]. A variety of physiological functions such as growth, metabolism, reproduction success, food consumption, and the capacity to maintain internal homeostasis capacity of aquatic species will be affected in response to temperature fluctuation [2].

Effects of heat stress on catfish industry

Catfish is the major aquaculture species in the United States. It is a temperate species with wide natural habitat and harbor great plasticity when they encounter temperature variations. It must undergo and adapt seasonal temperature changes from near freezing during winter in the North to over 36℃ in the summer in the South [5]. Water temperatures in aquaculture ponds currently approach upper thermal tolerance (~37℃) levels for channel catfish, particularly in June-August, which routinely see daily maximum values of 29℃ and higher [6]. Production rate in catfish ponds could be decreased duo to a decreased dissolved oxygen levels and an increased virulence of pathogens caused by temperature increases. A drop in dissolved oxygen levels can lower the management capacity of water from uneaten feed, fecal matter, and fish metabolism, which could lower the reproductive capacity of catfish ponds [7]. The immune function of catfish is compromised in stressful environment. On the one hand, the warmer water temperature can enhance the probabilities of certain disease and parasite outbreaks, especially in crowed catfish ponds. For instance, enteric septicemia of catfish (ESC) disease causes the largest economic loss in catfish industry. Pathogen of ESC, *Edwardsiella ictaluri*, presents highest virulence at 28-29°C; on the other hand, elevated body temperature can itself stress the fish and consequently make them vulnerable to be infected by pathogens [8].

Heat relationships to growth, reproduction, and survival of channel catfish have been recognized based on the evidence from physiological performance in catfish. Arnold et al. 2013 studied the effects of three cycling upper-range temperature regimes (23-27℃, 27-31°C, and 31-35°C) characteristic of aquaculture environments on juvenile channel catfish growth, feeding performance. The survival of catfish was significantly decreased for individuals in the warmest treatment compared with those in the coolest treatment. The growth of channel catfish

was decreased, which largely due to reduced food consumption and feed conversion and increased levels of activity. Therefore, increased water temperature, could likely caused by climate change, present challenges to the culture and management of catfish. In a 6-week growth experiment, catfish were subjected to daily cycling temperatures of either 27-31°C or 32-36°C, mimicking pond fluctuations. The results demonstrated the growth in weight and length of channel catfish decreases at temperatures greater than 27-31°C regardless of geographical origin of strain [6].

Heat tolerance between channel catfish, blue catfish and hybrid catfish

In United States, the majority of catfish production occurs in the southeast (92%), where some of the warmest conditions are found [9]. Channel catfish (*Ictalurus punctatus*) with the thermoplasticity has a natural geographical distribution from southern Canada to northern Mexico, which encompasses a temperature range from 5 to 35℃ [10, 11]. Blue Catfish (*Ictalurus furcatus*) distribute further south, ranging from the Mississippi River basin and Gulf Coast through Mexico and into Guatemala and Belize [12]. The inter-specific hybrid (channel catfish female × blue catfish male) were also expected to have a higher heat tolerance performance than channel catfish strains because blue catfish have a more southern distribution than channel catfish do [9]. It was reported that the optimum water temperature for channel catfish best growth performance ranging from 27-32°C. However, the Southeastern U.S. ponds reach daily maximum as high as of 34-36°C with daily fluctuations averaging 4°C in May-August [6]. Therefore, fluctuations of pond temperature and maximum daily water temperature are likely exacerbated by the global warming, which demonstrated that heat stress is a growing concern for pond culture of catfish.

Knowledge about the heat tolerance between different channel catfish strains, blue catfish strains and their hybrid catfish is very limited. It was first found that little to no geographic variation in incipient upper lethal temperature (IULT) of Channel Catfish from Florida and Ohio [13]. However, the study was limited by small sample sizes. Thereafter, critical thermal maximum (CT-max) was used to examine the thermal sensitivity of catfish to acute temperature fluctuations, which can provide guidelines for best culture management practices [10]. It was observed that the CT-max ranged from 38.6℃ to 40.3℃ for two geographically distinct strains of channel catfish by Stewart and Allen at 2014. Catfish with a southern distribution (Delta Select strain, from the Mississippi Delta, Mississippi) had a greater CT-max than did catfish with a northern distribution (Red River strain, from the Red River, North Dakota). These geographic differences in thermal tolerance were also observed in the hybrid catfish, suggesting a genetic component for thermal tolerance in catfish [9].

Enteric septicemia of catfish (ESC) disease

ESC disease and the effects on catfish industry

Catfish is the major aquaculture species in the United States, accounting for over 50% of all US aquaculture production. However, the catfish industry in the US has recently encountered unprecedented challenges due to devastating diseases and fierce international competition. In particular, enteric septicemia of catfish (ESC) disease, caused by a gram-negative bacteria pathogen, *E. ictaluri*, is one of the most serious infectious diseases in catfish and commonly associated with widespread mortality through both acute and chronic infections in ponds. ESC disease is temperature dependent with optimal temperatures of 24-28℃. As water temperature rises above 30℃ or drops below 20℃ the bacteria becomes inefficient and

dormant or dies. Therefore, it is a biphasic disease with annual occurrence in the spring and fall [14].

The disease was first identified and determined at Auburn University in 1976 [15]. After that, the disease has quickly throughout the industry, resulted the increasing from 8% of total disease cases in1979 to 30% of total disease cases in 1998 [16]. In 1999, economic losses due to ESC were estimated to be as high as $60 million. Until present, ESC is recognized as a strong economically significant pathogen in most fish stocks and culture ponds. ESC disease affects all size classes of catfish, develops rapidly, spreads easily and causes high mortality.

To date, there is no totally effective control applications for ESC disease. Although a live attenuated vaccine against *E.ictaluri* was developed, its effectiveness is only limited to a fraction of filed isolates [17]. In addition, the large numbers of fish make application of vaccine quite difficult. So far, injection is the most effective way of vaccine application; However, vaccination of a billion catfish through injection is extremely labor-intensive and adds to economic burden to catfish producers. Immersion is an alternative way with less laborious, but the results of vaccination to catfish by immersion vary from trail to trail. There are no effective antibiotics or therapeutics available for ESC. Based on these condition, improved brood stocks with high ESC disease resistance are desperately demanded. Genetic improvement of disease resistance is one of the most effective way to address disease problem in aquaculture. Understanding of disease controlling loci in catfish will guide the development of disease-resistant brood stocks. Therefore, marker-assisted selection or whole genome-based selection holds a great promise for supporting long-term sustainable aquaculture program.

Cause of ESC and clinic signs

Husbandry and environmental stress play significant roles in determining the clinical and economic impact of *E. ictaluri* infections. Nevertheless, this organism is considered a primary pathogen and is capable of causing substantial losses even on well-managed farms. Usually, ESC can occur when a susceptible host (channel catfish) encounters *E. ictaluri* under the right environmental conditions. Several stress factors such as handling, close confinement, improper diet, low water chlorides, poor water quality, and water temperature fluctuations can lead to increased susceptibility to infection. The introduction of ESC-infected fish into a pond containing healthy fish, or stocking healthy fingerlings into a pond containing older catfish that are carrying *E. ictaluri*, can result in the perpetuation and spread of ESC. Fish that survive an outbreak can carry the bacterium in the brain, kidney and liver for extended periods (up to 200 days). These survivors develop specific immunity that protects them from subsequent infection and disease [14].

E. ictaluri causes both physical and behavioral changes in catfish. A number of clinical signs are associated with ESC, including gradual appearance of small circular red spots over body, bloody areas on base of fins, white circular spots, raised reddish area on top of head, protruding eyes, ulcerated areas on top of the head (so called hole-in-the-head), bloated fluid filled belly, and heavy signs of hemorrhages of organs and tissues. Behavioral signs include reduction in feeding, erratic swimming, swirling, and hanging head up and tail down in the pond [14].

Variation ESC disease resistance between channel catfish, blue catfish and hybrid catfish

ESC disease resistance of catfish is usually heritable and variable among species and strains originating from different geographic locations (Dunham and Elaswad, 2018). Channel

catfish, the major cultivated species of the US catfish industry, is particularly sensitive to ESC infection. However, significant variation in channel catfish resistance was found among strains, families, and body weights. Red River strain channel catfish were the most resistant (14.9% mortality), followed by Mississippi-select fish (67.1%) and Mississippi-normal fish (72.3%). In contrast, blue catfish is more resistant to ESC than channel catfish [18-20]. An inter-specific hybrid catfish has been available for commercial use, and displays strong heterosis and superior performance traits including ESC disease resistance. Therefore, the hybrid catfish has become common in the catfish production industry in the United States accounting for 50-70% of all catfish produced [21]. Combination of strains and families of the parent species impacts the hybrid catfish disease resistance. For example, hybrids of Norris female channel catfish and blue catfish males had intermediate resistance between pure strain blue catfish and pure strain channel catfish [16, 22]. While hybrids of NWAC103 channel catfish and blue catfish had a better survival when exposed to natural diseases of ESC [23].

The molecular mechanisms underlying heat stress in catfish

Heat stress disturbs cellular homeostasis and can lead to severe retardation in growth and development, and even death. Knowledges from other model species reveal that a series of evolutionarily conserved stress-responsive genes display distinct expression for heat stress, including genes involved protein folding and repair, protein degradation and biosynthesis, energy metabolism, cell cycle and signaling, cytoskeletal reorganization and apoptosis [8, 24, 25]. Increasing the levels and magnitudes of stress sequentially can lead to different components of the stress resopnse[24-27]. For instance, under mild heat stress, chaperone

proteins could be induced to refold proteins that have unfolded caused by heat perturbation of tertiary structure, so that to maintain protein homeostasis [26]. At moderate levels of heat stress, the abnormal folded protein which cannot be rescued through activities of chaperones will be degraded by proteolysis through the ubiquitin-proteasome pathway. In addition, above a certain level of stress, basic cell activities such as cell proliferation and cytoskeletal reorganization may be induced because of the cellular damage, which has attendant effects on cellular structure and function. Especially to DNA, sufficient energy needs to be redirected from housekeeping functions toward the stress response. Furthermore, when suffering severe acute stress, significant enough damage to cell will to trigger induction of apoptotic pathways [25, 28].

Similarly, heat stress in catfish can compromise a variety of physiological functions including metabolism, fecundity and susceptibility of fish to disease or toxicants, which can even result in population-level effects. Catfish has great plasticity in dealing with environmental temperature variations considering it's wide geographic distribution that spans from North to South of United State. Therefore, catfish is not only an important aquaculture species but also a good research model for heat stress studies. In-deep analysis on molecular biology levels for response to thermal stress has long been of interest in catfish genetic and genomic research, but it remains incompletely documented.

In catfish, a number of heat shock genes were first characterized. Heat shock proteins are a class of proteins that are produced by cells in response to exposure to stressful conditions. Several heat shock proteins are play function as intra-cellular chaperons involved in the folding and unfolding of other proteins in response to heat, oxidative and other cellular stress [29]. The

heat-shock protein (stress-70 family) was isolated from channel catfish liver in 1994 [30]. It has been suggested that levels of synthesis or accumulation of Stress-70s may be useful in determining whether a particular environmcntal treatment is perceived by the organism as stressful. Straight after, the cDNA sequence of a member of the channel catfish heat shock protein 70 (CF HSP70) family was indentified, as well as expression in three leukocyte cell lines was determined in 1996 [31]. In this study, high levels of CF HSP70 mRNA were constitutively expressed at optimal culture temperature (27°C), whereas heat shock (37°C) elicited only a modest induction of CF HSP70 expression.

Transcriptom is one of the most rapid and versatile responses of organims expriencing environmental stress. In order to obtain a broad understanding of heat stress induced gene expression in catfish, a RNA-Seq analysis was conducted by using hybrid catfish, generated from crossing channel catfish female and blue catfish male, which is now widely used in aquaculture production because of its superior performance [8]. In this study, RNA-Seq was carried out on gill and liver samples from intolerant and tolerant catfish groups as well as from the control catfish group. A total of 2,260 unique genes showed significant differential expression between control fish and intolerant and/or tolerant fish in gill and/or liver after heat stress. After gene ontology, enrichment and pathway analysis, the differentially expressed genes were classified into six functional categories: protein folding, 2) protein degradation, 3) protein biosynthesis, 4) energy metabolism, 5) molecule and ion transport, and 6) cytoskeleton reorganization. Specifically, genes involved in oxygen transport, protein folding and degradation, and metabolic process were highly induced, while general protein synthesis was dramatically repressed in response to the lethal temperature stress. The most strongly inducible

genes in this RNA-Seq study were those of molecular chaperones, such as members of HSP40, HSP70 and HSP90 families, as well as cofactors. The chaperone proteins are critical in maintaining protein homeostasis durig cellular response to heat stress through interacting with denatured proteins, preventing their aggregation and degradation [32]. Some damaged proteins that can not enter the chaperone pathway are degradated by either autophagy-lysosomal pathway or ubiquitinproteasome pathway (UPP). A large number of proteases such as cathepsins and legumain were significanlt induced in this study. Genes encoding proteins involved in transporting various molecules and ions throughout the cell were identified after heat stress in catfish. Notebly, the gene products of transporting oxygen were most significantly upregulated in both gill and liver, such as several hemoglobin subunits. As expetced, the expression of genes involved in regulating metabolism and repair systerm showed up regulation in response to heat, bescause these processes are energy-costing. In contrast, several genes encoding enzymes involved in respiratory chain were repressed, including genes coding for mt-ND1, mt-ND2, mt-ND6, and cox2. It's well recognized that heat stress preferentially leads to upregulation of specific stress-related genes while downregulation of general genes involved in protein synthesis [28, 33]. The ribosomal protein genes were significantly repressed in gills of catfish exposed to high temperature. Besides the effects on internal cellular processes, heat stress can induce the expression of several cytoskeleton-associated proteins including Ras GTPase activating protein-binding protein 2 (G3BP2), contractile protein tropomyosin (TPM4), matrix metalloproteinase genes (MMP9, MMP13, AND MMP18) and collagen genes (COLLA1A and COLLA1B).

The molecular mechanisms underling ESC disease in catfish

To better understand the critical innate immune response of catfish with ESC disease, a large number of cytokines, CC chemokines, antimicrobial peptides, and Toll-like receptors were first identified and characterized through EST sequencing. For instance, highly duplicated nature of the catfish CC chemokine genes was identified, which allowed division of roles that may be manifested in spatial, temporal, or functional differences. With the inducible CC chemokines following *E. ictaluri* infection, such as SCYA105, SCYA109, SCYA112, SCYA113, SCYA115, SCYA117, and SCYA125, it is likely that they are involved in the attraction of leukocyte populations to the site of infection [34]. NK-lysin is an effector protein of cytotoxic T lymphocytes and natural killer cells. Three distinct NK-lysin transcripts exist in channel catfish, and exhibited distinct expression profiles in various tissues [35]. Hepcidin was initially isolated from human plasma ultrafiltrate and urine and shown to possess antimicrobial activities [36]. The channel catfish hepcidin gene was expressed in a wide range of tissues except brain, and induced after bacterial infection with *E.ictaluri*. The expression profile of the catfish hepcidin gene during the course of bacterial infection mirrors those of inflammatory proteins such as chemokines, suggesting an important role for hepcidin during inflammatory responses [37]. Toll-like receptors (TLRs) were the earliest characterized and the most extensively studied pathogen recognition receptors (PRRs). Twenty catfish TLR genes were extensively analyzed against their counterparts from various species, with tlr25 and tlr26 are tlrs identified only in channel catfish. After *E.ictaluri* infection, several TLR genes showed significantly up-regulated in the spleen and liver, but down-regulated in the head kidney, suggesting their involvement in the immune responses against the intracellular bacterial

pathogen in a tissue-specific manner in catfish, perhaps through rapid migration of phagocytes to infection sites [38].

Later on, high-density *in situ* oligonucleotide microarrays were developed to study the expression of these important immune components in the larger context of the catfish transcriptome following ESC infection. A Microarray analysis was conducted to evaluate the acute phase response (APR) in channel catfish liver following the *E. ictaluri* infection. The results revealed a well-developed APR in catfish, with particularly high upregulation of genes involved in iron homeostasis such as intelectin, hemopexin, haptoglobin, ferritin, and transferrin. Strong upregulation of the complement cascade, PRRs and chemokines indicated that the catfish liver likely plays an important role in pathogen recognition and defense as well as inflammatory signaling. [39]. Subsequently, microarray analysis of gene expression changes in blue catfish liver after *E. ictaluri* infection indicated strong upregulation of several pathways involved in the inflammatory immune response and potentially in innate disease resistance, which was similar to the observations in channel catfish. However, significant differences were noted between the two species at day 3 after infection including a set of 58 genes that were significantly differentially expressed in blue catfish but not in channel catfish. Notably, the early induction of several components of the MHC class I-related pathway following infection with an intracellular bacterium was first reported in catfish, perhaps providing an immune advantage [40]. Together, these microarray results add to our broad understanding of the immune responses following *E. ictaluri* infection, and provide potential insights into the molecular mechanisms for ESC disease.

Then, high-throughput RNA-Seq analyses were carried out to further characterize the

global gene expression profiles after *E. ictaluri* infection in catfish. Firstly, a RNA-seq was conducted to study the role of the intestinal epithelial barrier following *E. ictaluri* challenge. 1633 differentially expressed genes (DEGs) were identified from the comparison of digital gene expression between challenged and control samples at 3 h, 24 h, and 3 day following bacteria exposure. Gene pathway analysis of the DEGs revealed six broad functional categories reflective of likely cellular and physiological responses, with the centrality of actin cytoskeletal polymerization/remodelling and junctional regulation in pathogen entry and subsequent inflammatory responses. For instance, genes associated with bacterially-induced creation of actin-rich structures were observed to be differentially expressed in control and infected catfish intestine included Arp2/3, ezrin, filamin, Rho-GTPase, Cdc42SE2, integrins, gelsolin-like (capg), supervillin, and ahnak; transcripts representing junctional proteins were up-regulated, including cadherins, desmoplakin, and magi3. The other enriched functional categories were Lysosome/Phagosome, Immune Activation/Inflammation, Attachment/Pathogen Recognition, and Endocrine/Growth Disruption [41]. This study identified several novel expression patterns of teleost mucosal gene and highlighted unexpected roles for candidate genes and pathways often missed in a priori approaches. Subsequently, a bulked segregant RNA-Seq (BSR-Seq) approach using F2 backcross catfish progenies was conducted, with the objective to simultaneously study gene expression profiling and gene-associated SNPs for ESC resistance in catfish [42]. A total of 1,255 DEGs were identified between resistance and susceptible fish; 56,419 SNPs locating on 4,304 genes were found as significant SNPs between resistance and susceptible fish. Mapping of significant SNPs revealed 8 genomic regions that could involve in ESC disease resistance. Along with gene expression data, 17 DGEs including NLR, MHC-

related genes and Mannan-binding lectin serine peptidase were identified to be potentially candidate genes involved in ESC resistance. Importantly, this study demonstrated the use of BSR-Seq to analysis of genes and SNPs underling various performance traits in a cost-effective manner. The results laid a solid foundation for future functional characterization, genetic mapping, and QTL analysis of ESC resistant genes from catfish.

Genome-wide association study (GWAS)

Quantitative trait locus (QTL) mapping

Production traits are determined by composite of genes coupled with transcriptional regulation, post-transcriptional modification and regulation, translational regulation and post-translational modification and regulation, along with environmental impact and genotype-environment interactions [43]. A quantitative trait locus (QTL) is a section of genetic markers which correlates with variation in a phenotype. QTL mapping is a first step toward identification the phenotypic data (trait measurements) to genotypic data (usually molecular markers) to explain the genetic basis of variation underlying complex traits [44]. The ultimate goal of QTL mapping is usually to identify the causative genes and polymorphisms that contribute directly to control of the performance traits. Analysis of QTL responsible for complex traits is great of interest from the perspective of fundamental genetics and biology. In particular, results from QTL mapping can be applied in selective breeding programs by marker-assistant selection (MAS) or genomic selection (GS), with the objective to enhance the selection accuracy for many economic important traits.

Two main methodologies are well applied for QTL study, including linkage mapping and association mapping. Linkage mapping towards to identification of QTL based on their genetic

linkage to genetic marker within families. Associated mapping, also know as association study mapping, towards to detection QTLs based on the historically accumulated linkage disequilibrium (LD) between genetic markers and the traits of interest [45]. In order to conduct a QTL analysis efficiently, the selected genetic markers must have sufficiently high variation and polymorphisms among mapping families or populations. Various genetic markers have been applied in QTL mapping, including amplified fragment length polymorphism (AFLP), restriction fragment length polymorphism (RFLP), microsatellite (SSR) and single nucleotide polymorphism (SNP). At present, SNP markers are extensively applied in linkage mapping or association studies based on two main advantages: first, SNPs display high abundant and high density throughout the genomes of most specie, making SNPs adaptable to automation, and detect hidden polymorphism that can not be revealed by other markers [46]; second, SNPs are regraded biallelic marker in most cases and are inherited co-dominantly. With dramatical advancement of next-generation sequencing technologies, SNPs from high-throughput genotyping data are well available in a cost-effective way. With this respect, high-density SNP chips/arrays can be developed that offer high efficiency for genotyping and QTL mapping, such as the development of catfish 250K and 690K SNP arrays [47, 48], Atlantic salmon 132K array [49], rainbow trout 57K array [50] and carp 250K array [51].

GWAS analysis and its application in catfish

Genome-wide association study (GWAS) is to conduct association studies using genome-wide genotyping data. It allows detection of markers (SNPs) closely linked to QTLs, which based upon the principle of linkage disequilibrium (LD) between genetic markers and QTL that affect the trait [52]. Generally, loci that are physically close together exhibit stronger LD than

loci that are farther apart on a chromosome [53]. Moreover, GWAS can offer the chance or opportunity to systematically analyze the genetic structure of complex traits, and it can get benefits from the high diversity and rapid LD decay in a species [54, 55]. Although QTL mapping is well-suited for family-based samples, association studies, especially GWAS, do not rely on pedigree information and can potentially offer higher mapping resolution by genotyping population samples with high-density SNP arrays. Moreover, the rapid development of next-generation sequencing technologies and recent developments in GWAS methodologies have offered mature software packages for association analysis. The continuing decrease in the genotyping cost makes GWAS a standard tool for detecting natural variation that accounts for complex quantitative phenotypes in organisms [54]. During last decades, GWAS has evolved into a powerful tool for investigating the genetic architecture of important traits of human beings, crop, and livestock.

Economically important traits of catfish include disease resistance, growth rate, stress response, feed conversion efficiency, body conformation and processing yield, meat quality, tolerance to low dissolved oxygen, and tolerance to low water quality. With abundant and high quality genome resources, we developed catfish 250K and 690K SNP array using Affymetrix Axiom genotyping technology [47, 48]. The arrays provide valuable platform for GWAS, fine QTL mapping, high-density linkage map construction, haplotype analysis, and whole genome-based selection.

In catfish, GWAS have been successfully conducted to identify QTLs associated with several important performance traits, including disease resistance for columnaris [52] and ESC [56-58], growth rate[59], head size[60], body conformation[61], and low oxygen tolerance [62,

63]. For columnaris disease, a major QTL on linkage group 7 was identified with significantly associated with columnaris resistance/susceptibility. In addition, 3 additional suggestively associated QTL regions were identified on linkage groups 7, 12, and 14. Many candidate genes on the four associated regions are involved in PI3K pathway that is known to be required by many bacteria for efficient entry into the host [52]. For ESC, three GWAS that aimed to identify QTLs responsible for ESC resistance were conducted in channel catfish [58], the second and fourth generations of backcross catfish [56, 57]. The linkage groups of 1 and 26 were found to have significantly associated regions in channel catfish populations [58]; LG1 and LG23 were determined to be significantly associated with ESC resistance in second generation of hybrid catfish populations [56]; LG1, 12 and 26 16, 26 were revealed to be associated with ESC disease resistance in forth second generation of hybrid catfish populations [57]. Meanwhile, several positional candidate genes were identified form these GWAS studies including NKC1, FZD8, KLF6, NLRP12, ACBD5A, APBB1IP, MYO3A, STAT2, NLRC3, AGTR1, TRPC1, ABI1, ACTR3B, VAV3, MRC1L, PRKCQ and GATA3. For growth trait, a genomic region of approximately 1 Mb on linkage group 5 was found to be significantly associated with body weight. In addition, four suggestively associated QTL regions were identified on linkage groups 1, 2, 23 and 24. Most candidate genes in the associated regions are known to be involved in muscle growth and bone development, some of which were reported to be associated with obesity in humans and pigs, suggesting that the functions of these genes may be evolutionarily conserved in controlling growth [59]. All in all, these GWAS studies offered valuable chances and resources to systematically analyze the genetic structure of complex traits in catfish. Further fine mapping and functional studies should allow identification of the causal locus and

genes responsible for these performance traits in catfish, and facilitation of marker-assisted selection or whole genome-based selection for long-term sustainable aquaculture program.

Bulked segregant RNA-Seq (BSR-Seq)

RNA Sequencing (RNA-Seq)

RNA-Seq is an extremely efficient strategy on revealing global expression profiles under specific physiological conditions, development stages, or various environmental stimulations. The sequencing reads of any specific transcripts in the RNA library applied for RNA-Seq represent the levels of their presence in the RNA library, by which offering the corresponding gene expression levels using quantitative analysis. As such, it is widely used to determine differentially expressed genes and identify pathways controlling cell fate, development and disease progression in a host [41, 42, 64]. Also, it can be applied for genome survey, and reveal massive functional gene and molecular markers in a rapid and efficient manner [65].

RNA-Seq has superior advantages than microarray and tag-based transcriptome analyses, making it well applied to refresh our knowledge of eukaryotic transcriptome and address various biological questions. First and foremost, RNA-Seq has both qualitative and quantitative advantages, thereby, the expression levels of even low-abundance transcripts could be detected. Moreover, it allows identification and quantification of the expression levels of isoforms [8]. For non-model organisms of interest that reference genome assembly are unavailable, RNA-Seq can provide great opportunity to generate transcriptomes. Currently, the significant reduction of RNA-Seq cost enables a deep sequencing and better coverage than methods used previously, such as more technical and biological replicates of a single study can be simultaneously conducted. Apart from using polyA mRNA, which focus on the protein-coding

genes, RNA-Seq could be applied to identify total RNA, pre-mRNA, and various non-coding RNAs using various library construction methods [66]. Therefore, these advantages have made RNA-Seq a better choice for researchers who deal with transcriptomes.

Bulked segregant RNA-Seq (BSR-Seq)

Although RNA-Seq is widely used to determine differentially expressed genes under a specific "treatment", tts application used to be limited to gene expression profiling. However, through the utilization of specific families combined with elaborate experimental designs and SNP analysis, it is possible to analyze genetic segregation between treatments. One of these experimental designs is called bulk segregant RNA-Seq (BSR-Seq). RNA samples are collected from pooled individuals that fall into phenotype extremes, e.g., the best and the worst performers. Such pooled RNA samples are then subjected to RNA-Seq. The obtained RNA-Seq datasets can be analyzed not only for differentially expressed genes between the bulks (phenotypic extreme groups), but also for differences in allele usage. Strong association of differential usage of alleles in these phenotypic extreme bulks would suggest genetic linkage of the associated genes with the phenotype, whereby the positional candidate genes as well as the expression candidate genes can be identified [42, 67].

BSR-Seq is an integrated solution of bulk segregant analysis (BSA) and RNA-Seq technology. BSA was first characterizing in plants as to identify genetic markers related with specific genes or genomic regions using only two bulks of pooled DNA samples with contrasting phenotypes in disease resistance [68]. The basic idea of BSA is that phenotypic extremes (such as the resistant samples versus the susceptible samples) should have drastic differences in their genotypes. When samples are selected from phenotypic extremes, and their

corresponding genotypes are analyzed in each bulk, a correlation of genotypes with phenotypes should be expected. In other word, the variations among individuals may be quite subtle and difficult to detect; however, the pooled samples (bulk) of the phenotypic extremes should pose a strong contrast in their genotypes at the genomic location linked to the trait. This approach appeared to be flexible as different types of segregating populations can be used to generate extreme phenotypes and develop bulks [67].

The power of BSA analysis has been well correlated with the use of molecular markers, for markers are usually assumed to quantify the allelic frequencies that associated with the phenotypic bulks. At the beginning, BSA was employed with low through-put molecular markers including RFLPs [69], RAPDs [70, 71]. Then SSRs [72, 73] and amplified fragment length AFLPs [74] were involved in BSA study. Advances in NGS allowed effective discovery of massive molecular markers such as single-nucleotide polymorphism (SNPs), which promoted BSA in broader applications of NGS-based studies [75, 76]. The application of RNA-seq has accelerated gene expression profiling and identification of gene-associated SNPs in many species. Therefore, coupling with RNA-Seq, BSR-Seq has the full capability to identify differentially expressed genes (DEGs) between bulks, and also the ability to map significant SNPs to genes associated with the traits [42]. However, this approach has only been widely applied in several plants but still rarely employed in animals [77-81]. As pooling strategy is involved, BSR-Seq is well suited to species with high fecundities because it is easier to collect samples from phenotypic extremes with such species. As such, it should be extremely useful for aquaculture species.

In comparison, GWAS analysis utilizes SNP arrays to obtain statistical significant,

although very powerful, requires much greater financial resources. The restriction to allelic variants in the genome level limits their application when considering the allele-specific expression of candidate genes. BSR-Seq is a way more economically efficient. In addition, taking the example of disease resistant research, BSR-Seq also provide several layers of additional information including 1) Differentially expressed genes after treatment as compared with control fish, which would provide information of disease response and defense responses; 2) Differentially expressed genes between the resistant and the susceptible bulks, which may provide insights into the molecular basis of disease resistance and the gene pathways involved; 3) Cross analysis of expression candidates and positional candidates will allow identification of the differentially expressed genes within QTL regions, whereby candidate genes in the pool can be narrowed for future studies of causal genes. Therefore, BSR-Seq would be a powerful tool to map genes and QTLs underling economic traits of aquaculture species in a rapid and efficient manner.

References

1. Ficke AD, Myrick CA, Hansen LJ: **Potential impacts of global climate change on freshwater fisheries**. *Reviews in Fish Biology and Fisheries* 2007, **17**(4):581-613.

2. Pörtner H-O, Peck M: **Climate change effects on fishes and fisheries: towards a cause-and-effect understanding**. *Journal of Fish Biology* 2010, **77**(8):1745-1779.

3. Roessig JM, Woodley CM, Cech JJ, Hansen LJ: **Effects of global climate change on marine and estuarine fishes and fisheries**. *Reviews in fish biology and fisheries* 2004, **14**(2):251-275.

4. Todd AS, Coleman MA, Konowal AM, May MK, Johnson S, Vieira NK, Saunders JF: **Development of new water temperature criteria to protect Colorado's fisheries**. *Fisheries* 2008, **33**(9):433-443.

5. Ju Z, Dunham R, Liu Z: **Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation**. *Molecular Genetics and Genomics* 2002, **268**(1):87-95.

6. Arnold MB, Torrans EL, Allen PJ: **Influences of cyclic, high temperatures on juvenile channel catfish growth and feeding**. *North American Journal of Aquaculture* 2013, **75**(1):77-84.

7. Ficke AA: **The effects of global climate change on the fishes of the southeastern United States**. *WWF-World Wide Fund For Nature (formerly World Wildlife Fund), Gland, Switzerland* 2005.

8. Liu S, Wang X, Sun F, Zhang J, Feng J, Liu H, Rajendran K, Sun L, Zhang Y, Jiang Y: **RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish**. *Physiological genomics* 2013, **45**(12):462-476.

9. Stewart HA, Bosworth B, Petrie-Hanson L, Martin JA, Allen PJ: **Effects of chronic upper temperature regimes on growth of two geographical strains of channel and hybrid catfish**. *Aquaculture Research* 2014.

10. Bennett WA, McCauley RW, Beitinger TL: **Rates of gain and loss of heat tolerance in channel catfish**. *Transactions of the American Fisheries Society* 1998, **127**(6):1051-1058.

11. Tavares-Dias M, Moraes F: **Haematological and biochemical reference intervals for farmed channel catfish**. *Journal of Fish Biology* 2007, **71**(2):383-388.

12. Graham K, Irwin E, Hubert W, Rabeni C, Schramm Jr H, Coon T: **A review of the biology and management of blue catfish**. In: *Catfish 2000: proceedings of the international ictalurid symposium American Fisheries Society, Symposium: 1999*; 1999:

37-49.

13. Hart JS: **Geographic variations of some physiological and morphological characters in certain freshwater fish.** . *Publications of the Ontario Fisheries Research Laboratory* 1952, **72**:1-79.

14. Chappell J: **Enteric Septicemia of Catfish**. 2008.

15. Hawke JP, McWHORTER AC, STEIGERWALT AG, BRENNER DJ: ***Edwardsiella ictaluri* sp. nov., the causative agent of enteric septicemia of catfish**. *International Journal of Systematic and Evolutionary Microbiology* 1981, **31**(4):396-400.

16. Hawke JP, Durborow R, Thune R, Camus A: **ESC: enteric septicemia of catfish**, vol. 477: Southern Regional Aquaculture Center Stoneville, Mississippi; 1998.

17. Shoemaker CA, Klesius PH, Bricker JM: **Efficacy of a modified live *Edwardsiella ictaluri* vaccine in channel catfish as young as seven days post hatch**. *Aquaculture* 1999, **176**(3-4):189-193.

18. Wolters WR, Johnson MR: ***Enteric septicemia* resistance in blue catfish and three channel catfish strains**. *Journal of Aquatic Animal Health* 1994, **6**(4):329-334.

19. Wolters WR, Wise DJ, Klesius PH: **Survival and antibody response of channel catfish, blue catfish, and channel catfish female × blue catfish male hybrids after exposure to *Edwardsiella ictaluri***. *Journal of Aquatic Animal Health* 1996, **8**(3):249-254.

20. Tucker CS, Hargreaves JA: **Biology and culture of channel catfish**, vol. 34: Elsevier; 2004.

21. Dunham RA, Elaswad A: **Catfish Biology and Farming**. *Annual review of animal biosciences* 2017(0).

22. Wolters WR, Wise DJ, Klesius PH: **Survival and antibody response of channel catfish, blue catfish, and channel catfish female× blue catfish male hybrids after exposure to *Edwardsiella ictaluri***. *Journal of Aquatic Animal Health* 1996, **8**(3):249-254.

23. Dunham RA, Umali GM, Beam R, Kristanto AH, Trask M: **Comparison of production traits of NWAC103 channel catfish, NWAC103 channel catfish× blue catfish hybrids, Kansas Select 21 channel catfish, and blue catfish grown at commercial densities and exposed to natural bacterial epizootics**. *North American Journal of Aquaculture* 2008, **70**(1):98-106.

24. Kültz D: **Molecular and evolutionary basis of the cellular stress response**. *Annu Rev Physiol* 2005, **67**:225-257.

25. Logan CA, Somero GN: **Effects of thermal acclimation on transcriptional responses to acute heat stress in the eurythermal fish *Gillichthys mirabilis* (Cooper)**. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2011, **300**(6):R1373-R1383.

26. Feder ME, Hofmann GE: **Heat-shock proteins, molecular chaperones, and the stress response: evolutionary and ecological physiology**. *Annual review of physiology* 1999, **61**(1):243-282.

27. Logan CA, Buckley BA: **Transcriptomic responses to environmental temperature in eurythermal and stenothermal fishes**. *The Journal of experimental biology* 2015, **218**(12):1915-1924.

28. Buckley BA, Gracey AY, Somero GN: **The cellular response to heat stress in the goby *Gillichthys mirabilis*: a cDNA microarray and protein-level analysis**. *Journal of Experimental Biology* 2006, **209**(14):2660-2677.

29. Buckley BA, Hofmann GE: **Magnitude and duration of thermal stress determine kinetics of hsp gene regulation in the goby *Gillichthys mirabilis***. *Physiological and biochemical zoology* 2004, **77**(4):570-581.

30. Abukhalaf IK, Zimmerman EG, Dickson KL, Masaracchia RA, Donahue MJ, Covington S: **Purification of the 70-kDa heat-shock protein from catfish liver: Immunological comparison of the protein in different fish species and its potential use as a stress indicator**. *Environmental toxicology and chemistry* 1994, **13**(8):1251-1257.

31. Luft JC, Wilson MR, Bly JE, Miller NW, Clem LW: **Identification and characterization of a heat shock protein 70 family member in channel catfish (*Ictalurus punctatus*)**. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 1996, **113**(1):169-174.

32. Parsell D, Lindquist S: **The function of heat-shock proteins in stress tolerance: degradation and reactivation of damaged proteins**. *Annual review of genetics* 1993, **27**(1):437-496.

33. Buckley BA, Hofmann GE: **Thermal acclimation changes DNA-binding activity of heat shock factor 1 (HSF1) in the goby *Gillichthys mirabilis*: implications for plasticity in the heat-shock response in natural populations**. *Journal of Experimental Biology* 2002, **205**(20):3231-3240.

34. Peatman E, Bao B, Peng X, Baoprasertkul P, Brady Y, Liu Z: **Catfish CC chemokines: genomic clustering, duplications, and expression after bacterial infection with *Edwardsiella ictaluri***. *Molecular Genetics and Genomics* 2006, **275**(3):297-309.

35. Wang Q, Wang Y, Xu P, Liu Z: **NK-lysin of channel catfish: gene triplication,**

sequence variation, and expression analysis. *Molecular immunology* 2006, **43**(10):1676-1686.

36. Krause A, Neitz S, Mägert H-J, Schulz A, Forssmann W-G, Schulz-Knappe P, Adermann K: **LEAP-1, a novel highly disulfide-bonded human peptide, exhibits antimicrobial activity**. *FEBS letters* 2000, **480**(2-3):147-150.

37. Bao B, Peatman E, Li P, He C, Liu Z: **Catfish hepcidin gene is expressed in a wide range of tissues and exhibits tissue-specific upregulation after bacterial infection**. *Developmental & Comparative Immunology* 2005, **29**(11):939-950.

38. Zhang J, Liu S, Rajendran K, Sun L, Zhang Y, Sun F, Kucuktas H, Liu H, Liu Z: **Pathogen recognition receptors in channel catfish: III Phylogeny and expression analysis of Toll-like receptors**. *Developmental & Comparative Immunology* 2013, **40**(2):185-194.

39. Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang S, Somridhivej B, Dunham R: **Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium**. *Developmental & Comparative Immunology* 2007, **31**(11):1183-1196.

40. Peatman E, Terhune J, Baoprasertkul P, Xu P, Nandi S, Wang S, Somridhivej B, Kucuktas H, Li P, Dunham R: **Microarray analysis of gene expression in the blue catfish liver reveals early activation of the MHC class I pathway after infection with *Edwardsiella ictaluri***. *Molecular immunology* 2008, **45**(2):553-566.

41. Li C, Zhang Y, Wang R, Lu J, Nandi S, Mohanty S, Terhune J, Liu Z, Peatman E: **RNA-seq analysis of mucosal immune responses reveals signatures of intestinal barrier disruption and pathogen entry following *Edwardsiella ictaluri* infection in channel catfish, *Ictalurus punctatus***. *Fish & shellfish immunology* 2012, **32**(5):816-827.

42. Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, Song L, Feng J, Liu S, Liu Z: **Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish**. *BMC genomics* 2013, **14**(1):929.

43. Geng X, Zhi D, Liu Z: **Genome-wide association studies of performance traits**. *Bioinformatics in Aquaculture: Principles and Methods* 2017:415-433.

44. Miles C, Wayne M: **Quantitative trait locus (QTL) analysis**. *Nature Education* 2008, **1**(1):208.

45. Gutierrez AP, Houston RD: **Quantitative Trait locus mapping in aquaculture species: principles and practice**. *Bioinformatics in Aquaculture: Principles and Methods* 2017:392-414.

46. Liu Z: **Single nucleotide polymorphism (SNP)**. In: *Aquaculture Genome Technologies*. Blackwell USA; 2007: 59-72.

47. Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y, Zhang J, Feng J, Kaltenboeck L, Kucuktas H: **Development of the catfish 250K SNP array for genome-wide association studies**. *BMC research notes* 2014, **7**(1):135.

48. Zeng Q, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S, Yang Y, Bao L, Yuan Z, Li N: **Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence**. *Scientific reports* 2017, **7**:40347.

49. Houston RD, Taggart JB, Cézard T, Bekaert M, Lowe NR, Downing A, Talbot R, Bishop SC, Archibald AL, Bron JE: **Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*)**. *BMC genomics* 2014, **15**(1):90.

50. Palti Y, Gao G, Liu S, Kent M, Lien S, Miller M, Rexroad C, Moen T: **The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout**. *Molecular Ecology Resources* 2015, **15**(3):662-672.

51. Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y, Kuang Y, Zhang Y, Feng J, Li C: **Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*)**. *Bmc Genomics* 2014, **15**(1):307.

52. Geng X, Sha J, Liu S, Bao L, Zhang J, Wang R, Yao J, Li C, Feng J, Sun F: **A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance**. *BMC genomics* 2015, **16**(1):196.

53. Visscher PM, Brown MA, McCarthy MI, Yang J: **Five years of GWAS discovery**. *The American Journal of Human Genetics* 2012, **90**(1):7-24.

54. Li H, Peng Z, Yang X, Wang W, Fu J, Wang J, Han Y, Chai Y, Guo T, Yang N: **Genome-wide association study dissects the genetic architecture of oil biosynthesis in maize kernels**. *Nature Genetics* 2013, **45**(1):43-50.

55. Yan J, Warburton M, Crouch J: **Association mapping for enhancing maize (L.) genetic improvement**. *Crop science* 2011, **51**(2):433-449.

56. Zhou T, Liu S, Geng X, Jin Y, Jiang C, Bao L, Yao J, Zhang Y, Zhang J, Sun L: **GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance**. *Molecular Genetics and Genomics* 2017, **292**(1):231-242.

57. Tan S, Zhou T, Wang W, Jin Y, Wang X, Geng X, al el: **GWAS analysis of QTL for**

resistance against *Edwardsiella ictaluri* in F2 interspecific hybrid catfish. *Molecular Genetics and Genomics* 2018: 1-14.

58. Shi H, Zhou T, Wang X, Yang Y, Wu C, Liu S, al el: **Identification of QTL and candidate genes associated with the resistance for enteric septicemia of catfish using genome-wide association analysis in channel catfish.** *Molecular Genetics and Genomics* 2018, **(Accepted)**.

59. Li N, Zhou T, Geng X, Jin Y, Wang X, Liu S, Xu X, Gao D, Li Q, Liu Z: **Identification of novel genes significantly affecting growth in catfish through GWAS analysis**. *Molecular Genetics and Genomics* 2017:1-13.

60. Geng X, Liu S, Yao J, Bao L, Zhang J, Li C, Wang R, Sha J, Zeng P, Zhi D: **A genome-wide association study identifies multiple regions associated with head size in catfish**. *G3: genes, genomes, genetics* 2016, **6**(10):3389-3398.

61. Geng X, Liu S, Yuan Z, Jiang Y, Zhi D, Liu Z: **A genome-wide association study reveals that genes with functions for bone development are associated with body conformation in catfish**. *Marine Biotechnology* 2017, **19**(6):570-578.

62. Zhong X, Wang X, Zhou T, Jin Y, Tan S, Jiang C, Geng X, Li N, Shi H, Zeng Q: **Genome-wide association study reveals multiple novel QTL associated with low oxygen tolerance in hybrid catfish**. *Marine Biotechnology* 2017, **19**(4):379-390.

63. Wang X, Liu S, Jiang C, Geng X, Zhou T, Li N, Bao L, Li Y, Yao J, Yang Y: **Multiple across-strain and within-strain QTLs suggest highly complex genetic architecture for hypoxia tolerance in channel catfish**. *Molecular Genetics and Genomics* 2017, **292**(1):63-76.

64. Sun L, Liu S, Bao L, Li Y, Feng J, Liu Z: **Claudin multigene family in channel catfish and their expression profiles in response to bacterial infection and hypoxia as revealed by meta-analysis of RNA-Seq datasets**. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* 2015, **13**:60-69.

65. Sudhagar A, Kumar G, El-Matbouli M: **Transcriptome analysis based on rna-seq in understanding pathogenic mechanisms of diseases and the immune system of fish: A comprehensive review**. *International journal of molecular sciences* 2018, **19**(1):245.

66. Kukurba KR, Montgomery SB: **RNA sequencing and analysis**. *Cold Spring Harbor protocols* 2015, **2015**(11):pdb. top084970.

67. Yao J, Wang R, Liu Z: **Genetic analysis using rna‐seq: bulk segregant rna‐seq**. *Bioinformatics in Aquaculture: Principles and Methods* 2017:169-178.

68. Michelmore RW, Paran I, Kesseli R: **Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers**

in specific genomic regions by using segregating populations. *Proceedings of the national academy of sciences* 1991, **88**(21):9828-9832.

69. Lister C, Dean C: **Recombinant inbred lines for mapping RFLP and phenotypic markers in *Arabidopsis thaliana***. *The Plant Journal* 1993, **4**(4):745-750.

70. Hormaza J, Dollo L, Polito V: **Identification of a RAPD marker linked to sex determination in *Pistacia vera* using bulked segregant analysis**. *Theoretical and applied Genetics* 1994, **89**(1):9-13.

71. Guangli S, Zhanjing H, Congfen H, Yinzhu S, Wang J: **Identification of the molecular markers linked to the salt-resistance locus in the wheat using RAPD-BSA technique**. *Acta Botanica Sinica* 2001, **43**(6):598-602.

72. Cheema KK, Grewal NK, Vikal Y, Sharma R, Lore JS, Das A, Bhatia D, Mahajan R, Gupta V, Bharaj TS: **A novel bacterial blight resistance gene from *Oryza nivara* mapped to 38 kb region on chromosome 4L and transferred to *Oryza sativa* L**. *Genetics Research* 2008, **90**(5):397-407.

73. Dussle C, Quint M, Melchinger A, Xu M, Lübberstedt T: **Saturation of two chromosome regions conferring resistance to SCMV with SSR and AFLP markers by targeted BSA**. *Theoretical and Applied Genetics* 2003, **106**(3):485-493.

74. Asnaghi C, Roques D, Ruffel S, Kaye C, Hoarau J-Y, Télismart H, Girard J-C, Raboin L-M, Risterucci A-M, Grivet L: **Targeted mapping of a sugarcane rust resistance gene (Bru1) using bulked segregant analysis and AFLP markers**. *Theoretical and Applied Genetics* 2004, **108**(4):759-764.

75. Miller AC, Obholzer ND, Shah AN, Megason SG, Moens CB: **RNA-seq–based mapping and candidate identification of mutations from forward genetic screens**. *Genome research* 2013, **23**(4):679-686.

76. Becker A, Chao D-Y, Zhang X, Salt DE, Baxter I: **Bulk segregant analysis using single nucleotide polymorphism microarrays**. *PLoS One* 2011, **6**(1):e15993.

77. Liu S, Yeh C-T, Tang HM, Nettleton D, Schnable PS: **Gene mapping via bulked segregant RNA-Seq (BSR-Seq)**. *PloS one* 2012, **7**(5):e36406.

78. Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M, Uauy C: **Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat**. *BMC Plant Biology* 2012, **12**(1):14.

79. Yates SA, Swain MT, Hegarty MJ, Chernukin I, Lowe M, Allison GG, Ruttink T, Abberton MT, Jenkins G, Skøt L: **De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification**. *BMC genomics* 2014, **15**(1):453.

80. Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, Berry S, Jack P, Caccamo M, Uauy C: **RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat**. *Plant biotechnology journal* 2015, **13**(5):613-624.

81. Du H, Zhu J, Su H, Huang M, Wang H, Ding S, Zhang B, Luo A, Wei S, Tian X: **Bulked segregant RNA-seq reveals differential expression and SNPs of candidate genes associated with waterlogging tolerance in maize**. *Frontiers in plant science* 2017, **8**:1022.

Chapter 2. Genome wide associate study (GWAS) for heat stress-associated SNPs in catfish

Abstract

Heat tolerance is a complex and economically important trait for catfish genetic breeding programs. With global climate change, it is becoming an increasingly important trait. To better understand the molecular basis of heat stress, a genome-wide association study (GWAS) was carried out using the 250 K catfish SNP array with interspecific backcross progenies, which derived from crossing female channel catfish with male F1 hybrid catfish (female channel catfish × male blue catfish). Three significant associated SNPs were detected by performing an EMMAX approach for GWAS. The SNP located on linkage group 14 explained 12.1% of phenotypical variation. The other two SNPs located on linkage group 16 explained 11.3% and 11.5% of phenotypical variation, respectively. A total of 15 genes with heat stress related functions were detected within the significant associated regions. Among them, 5 genes have known function in protein degradation process through ubiquitination pathway, including TRAF2, FBXW5, ANAPC2, UBR1 and KLHL29. Other genes conferring to heat stress include genes involved in protein biosynthesis (PRPF4 and SYNCRIP), protein folding (DNAJC25), molecule and iron transport (SLC25A46 and CLIC5), cytoskeletal reorganization (COL12A1) and energy metabolism (COX7A2, PLCB1 and PLCB4) processes. The results provide fundamental information of genes and pathways for further investigation on the molecular mechanisms of heat stress. The associated SNPs could be promising candidates for selecting heat-tolerant catfish lines after validating their effects on larger and various catfish populations.

Introduction

Catfish is one of the top agricultural commodities, and pivotal to employment opportunities in rural areas of the southeastern United States. Ongoing global climate change will lead to continuous rise in temperature, which may become a major stressor for fishes living in natural or artificial systems [1]. Therefore, developing heat-tolerant catfish lines becomes an important goal for genetic breeding programs through using various approaches, such as strain selection, crossbreeding, and hybridization. The hybrid catfish is expected to have a higher heat tolerance performance than channel catfish (*I. punctatus*) because blue catfish (*I. furcatus*) has a more southern distribution than channel catfish [2]. As a species of ectotherms, catfish must undergo and adapt seasonal temperature changes, ranging from near freezing during winter in the North to over 36 ℃ in the earthen ponds during the summer in the Southeastern U.S. [3, 4]. The shift of ambient water temperature can directly influence and/or disturb a variety of physiological functions of catfish. Thus, catfish, as a major aquaculture species, can also serve as a model species for heat stress studies [5].

GWAS allow the detection of linked QTLs in the families as well as those historically accumulated recombination events. F2 generation of hybrids produced by backcrossing design along with the highly segregated phenotypes provides a useful system for QTL analysis [6]. With a reference genome, candidate genes physically close to QTLs can be detected, which is useful for underlying biology of a trait by identifying in proximity to QTL [7, 8]. Results generated from GWAS can facilitate the selection of breeds and species resistant to heat stress [9]. However, no genome wide research on QTLs for heat stress was conducted in aquaculture species. Here, we conducted a genome-wide scan for QTLs conferring resistance to heat stress

using interspecific backcross progenies and the 250 K catfish SNP array, with the objective of initial understanding of the genomic regions important for heat stress in catfish.

Materials and methods

Heat stress challenge

A total of 630 catfish generated from two back cross families by crossing male F1 hybrid catfish (female channel catfish × male blue catfish) with female channel catfish were used for heat stress experiment. Each family containing 315 catfish progenies was transferred to an experimental tank for heat stress treatment after rearing for 2 weeks prior to challenge. The heat stress treatment followed the procedures conducted by Liu et al (2013). Briefly, all the fishes were acclimated for 72 hours at ambient temperature (24℃) before experiment for each family. A pace of 1℃/h of water temperature was increased for the experimental fish until it reached 36 ℃. Then, the temperature was held constant at 36 ℃, and the fish were closely monitored for signs of stress. The first fish showing lost equilibrium (LOE) was observed after 3 hours at 36 ℃ in both of the two families, and the last fish showing LOE was observed about 55 hours thereafter. The first and last 48 fish individuals lost balance in each family were continuously removed from the tank and sampled, which represented the most heat sensitive group and most heat tolerant group respectively.

Preparation of genomic DNA and genotyping

DNA from selected blood samples were isolated using standard protocols. Briefly, the blood samples were incubated at 55°C overnight and DNA were extracted twice with phenol and once with chloroform. DNA were precipitated by isopropanol and collected by brief

centrifugation, washed twice with 70% ethanol, air-dried, and resuspended in reduced EDTA TE buffer (10 mM Tris–HCl, 0.1 mM EDTA, pH 8.0). DNA were quantified using spectroscopy by Nanodrop (Thermo Scientific). The integrity of DNA samples then checked by 1% agarose gel electrophoresis stained with ethidium bromide. Then, DNA were diluted to 50 ng/uL and the quality of genomic DNA satisfied the requirement for the genotyping platform of SNP arrays. Genotyping using the catfish 250 K SNP array were outsourced and performed at GeneSeek (Lincoln, Nebraska, USA).

GWAS analysis

GWAS analysis was primarily undertaken using SNP & Variation Suit 8.0 (SVS8) software package (Golden Helix Inc.). We applied predetermined quality control metrics to the GWAS data. The analyses were restricted to samples for which > 90% of SNPs were successfully genotyped. In each sample series, we exclude SNPs with a calling rate of < 95% and MAF < 5%. For genome-wide association analysis, population stratification was first assessed using PCA of genotyped SNPs after removal of those within long-range LD regions and without further LD pruning. Association test was carried out by using EMMAX (Effieient Mixed-Model Association eXpedited) analyses in SVS. EMMAX analysis was conducted by using all informative SNPs with the first two principle components and the fish body weight as covariates to screen for genome-wide association. The model is listed as follows:

$$Y = X\beta + Zu + e$$

where Y is the vector of phenotype (0/1 as sensitive/tolerant); X is a matrix of fixed effects; β is the coefficient vector of fixed effects including first three principal components and fish body weight; Z is a matrix of random additive genetic effects; u is the vector of the random effect,

$Var(u) = G\sigma_g^2$, where σ_g^2 is the additive genetic variance and G is the genomic kinship matrix using the IBS; e is the vector of random residuals [6, 10].

Threshold P-value for genome-wide significance was determined based on 5% Bonferroni-correction with the estimated number of LD pruned SNP markers. A Manhattan plots of the P-value were produced using the SVS software, with the genetic marker map according to channel catfish reference genome sequence [11].

Sequence analysis

Genes within ±1 Mb of the significant SNPs associated with heat stress were predicted from the catfish genome sequences by FGENESH [12] and annotated by BLAST analysis and non-redundant (nr) protein database [13].

Results

Experimental fish and sample structure

Overall, 192 catfish backcross progenies (average body weight is 502.9 gram) were selected from the extremes of tolerance capability of the 630 fish based on the selective genotyping method. With known family pedigree information, principle component analysis was conducted using eigenvalues as coordinates to visualize the sample structure. As shown in Figure 1, Family A and B were actually two families and unrelated.

Quality control and linkage disequilibrium (LD) analysis

200,584 SNPs were kept in the subsequent GWAS after filtering out SNPs with an inheritance or genotyping error, a minor allele frequency (MAF) < 5%, or call rate < 95%. LD pruning was conducted with a window size of 50 SNPs, a step of 5 SNPs, and R2 threshold of 0.5. Threshold P-value for genome-wide significance was determined based on 5% Bonferroni-

correction with the estimated number of 8,091 independent markers. Thus, the genome-wide significance threshold was 0.05/8091=6.180 × e-6 (-log10 (P-value) = 5.209).

SNPs associated with heat stress

The GWAS identified three significant SNP markers conferring response to heat stress at the genome-wide significant level (-log10(P-value) > 5.209) (Figure 2). One SNP (AX-85318076) located on linkage group 14, the other two SNP (AX-86115098 and AX-85265807) located on linkage group 16. As shown in Table 1, their minor allele frequency (MAF) vary from 0.153 to 0.322, and the ratio of phenotypic variation (R2) explained by the SNPs vary from 11.3% to 12.1%. The nearest gene to the SNP in linkage group 14 is leukemia inhibitory factor receptor (LIFR), which is a signal-transducing molecule. SNP AX-86115098 in linkage group 16 is located within the collagen alpha-1(XII) chain (COL12A1) gene. COL12A1 plays function role in skeletal structure development as extracellular matrix structural constituent conferring tensile strength. The gene 1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1 (PLCB1) is identified nearby the SNP AX-85265807. PLCB1 has phosphatidylinositol phospholipase C activity to modify the composition of lipid bilayers.

Genes located within the associated QTL region underling heat stress

In this study, in additional to mapping the heat related SNPs, we sought to explore the candidate genes in the catfish genome sequence [11] of 1-Mb windows (SNP position ± 0.5 Mb) surrounding each identified promising SNP. A total of 15 candidate genes were determined and listed in Table 2, with their corresponding genomic positions and biological processes related with heat stress.

Discussion

In this study, we conducted a genome-wide scan for QTLs conferring resistance to heat stress using interspecific backcross progenies and the 250 K catfish SNP array [14], with the objective of initial understanding of the genomic regions important for heat stress in catfish. Three significant associated SNPs were detected by performing an EMMAX approach for GWAS. A total of 15 genes with heat stress related functions were detected within the significant associated regions. The results provide fundamental information of genes and pathways for further investigation on the molecular mechanisms of heat stress. The associated SNPs could be promising candidates for selecting heat-tolerant catfish lines after validating their effects on larger and various catfish populations.

In the present study, EMMAX was utilized for correcting population stratification which can lead to biased or spurious results. The statistical approach of EMMAX uses high-density markers to calculate a pairwise relatedness matrix representing the sample structure and correct for the structure during the mapping [10]. To exclude false positive results produced by sample structure observed in our study, EMMAX method was applied and adjusted for the first three principal components after calculation of kinship matrix-pairwise IBS distance. EMMAX has been widely applied in association study and yields more accurate statistics than other methods, such as GRAMMAR (genome-wide rapid association using mixed model and regression) [15].

The QTLs identified in this study explained a limited fraction of the phenotypic variance of heat tolerance. First of all, limited sample size and the population specificity of QTLs is the most important reason why our family-based association mapping cannot detect all the QTLs

associated with heat stress [16]. Various families showed dramatically different sensibility to heat stress, even though they are within the same strain, [17]. The use of two families allowed us to detect the QTLs, but provided a restricted power by taking advantage of historically accumulated recombination. Secondly, segregating alleles within one species could lead to decreased power of analysis, especially in the case that one parental species systematically carries tolerance alleles while the other one carries sensibility alleles [18]. Therefore, the region cannot be detected with strong significance using intraspecific SNPs. Thirdly, extremely complex genetic architecture of tolerance for heat stress could exist in catfish. Multiple QTLs with moderate or minor effects are hard to detected. Lastly, but not leastly, allele variations at the genome level only accounts for only a fraction of phenotypic variations. Gene expression regulations at multiple levels such as transcriptional and posttranscriptional levels, as well as environment and genotype-environment interactions can have a profound impact on the final phenotype in performance traits [6].

In catfish, response to heat stress has been reported related with several physiological and gene pathways by RNA-Seq analysis. For instance, genes involved in protein folding and degradation, protein biosynthesis, and energy metabolic process were highly induced under lethal temperature [5]. In general, the results demonstrated that complex molecular mechanisms were involved in heat tolerance other than simply induction of certain category of genes. Of the 15 genes identified in this study (Table 2), 5 genes were found to have known function in protein degradation process, including TNF receptor-associated factor 2 (TRAF2), F-box/WD repeat-containing protein 5 (FBXW5), anaphase-promoting complex subunit 2 (ANAPC2), E3 ubiquitin-protein ligase UBR1 (UBR1), kelch-like protein 29 (KLHL29). Interestingly, all of

them are involved in the protein ubiquitination pathway. The protein ubiquitination pathway has been reported to play a crucial role in response to heat stress in catfish [5], goby fish [19, 20], Arctic charr [21], and bluefin tuna [22]. It may suggest that the heat stress caused abnormal fold and irreversible damage to proteins which unable to enter the molecular chaperone pathway [20, 23]. In order to avoid forming cytotoxic aggregates, such damaged proteins need to be removed via proteolytic degradation by covalently tagged with multiple units of ubiquitin, when conjugated to a damaged polypeptide [24]. Therefore, there is increasing necessity of degradation for cells which are suffering sufficient levels of protein damage under such lethal heat treatment.

In addition to the needs for enhanced protein degradation proteins, cellular response to heat stress involves a range of biological mechanisms in order to stabilize cellular function, such as inhibition of DNA and protein synthesis, cell cycle arrest, molecule and iron transport, cytoskeleton reorganization and increased apoptosis [5, 25, 26]. Here, we also find genes participating in these biological processes and pathways in the genome-wide significant associated regions (Table 2). It includes genes involved in protein biosynthesis (PRPF4 and SYNCRIP), protein folding (DNAJC25), molecule and iron transport (SLC25A46 and CLIC5), cytoskeletal reorganization (COL12A1), and energy metabolism (COX7A2, PLCB1 and PLCB4). In aquaculture species, heat stress has been proved to disturb cellular homeostasis and can lead to severe retardation in growth and development, or even death. Taken together, the 15 genes detected in this study are involved in these biological processes, suggesting their importance for heat stress response in catfish.

This is the first association analysis in aquaculture species at the whole genome level to

investigate the genomic locus and genes related to heat stress. The results provide a valuable base of genes and pathways to further investigate their possible functions in heat stress. Considering the population specificity of QTLs and minor allele effect in association analysis, using larger or more catfish families, and various catfish strains are fairly necessary for fine mapping and accurate GWAS for heat stress analysis in the future.

Acknowledgements

This project was supported by a competitive grant (no. 2015-67015-22907) from the Tools and Resources-Animal Breeding, Genetics and Genomics Program of the USDA National Institute of Food and Agriculture (NIFA), and partially supported by USDA Aquaculture Research Program Competitive Grant no. 2014-70007-22395.

References

1. Ficke AD, Myrick CA, Hansen LJ: **Potential impacts of global climate change on freshwater fisheries**. *Reviews in Fish Biology and Fisheries* 2007, **17**(4):581-613.

2. Stewart HA, Bosworth B, Petrie-Hanson L, Martin JA, Allen PJ: **Effects of chronic upper temperature regimes on growth of two geographical strains of channel and hybrid catfish**. *Aquaculture Research* 2014.

3. Ju Z, Dunham R, Liu Z: **Differential gene expression in the brain of channel catfish (*Ictalurus punctatus*) in response to cold acclimation**. *Molecular Genetics and Genomics* 2002, **268**(1):87-95.

4. Arnold MB, Torrans EL, Allen PJ: **Influences of cyclic, high temperatures on juvenile channel catfish growth and feeding**. *North American Journal of Aquaculture* 2013, **75**(1):77-84.

5. Liu S, Wang X, Sun F, Zhang J, Feng J, Liu H, Rajendran K, Sun L, Zhang Y, Jiang Y: **RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish**. *Physiological genomics* 2013, **45**(12):462-476.

6. Geng X, Sha J, Liu S, Bao L, Zhang J, Wang R, Yao J, Li C, Feng J, Sun F: **A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance**. *BMC genomics* 2015, **16**(1):196.

7. Cole JB, Wiggans GR, Ma L, Sonstegard TS, Lawlor TJ, Crooker BA, Van Tassell CP, Yang J, Wang S, Matukumalli LK: **Genome-wide association analysis of thirty one production, health, reproduction and body conformation traits in contemporary US Holstein cows**. *BMC genomics* 2011, **12**(1):408.

8. Dikmen S, Cole JB, Null DJ, Hansen PJ: **Genome-wide association mapping for identification of quantitative trait loci for rectal temperature during heat stress in Holstein cattle**. *PLoS One* 2013, **8**(7):e69202.

9. Nayan V, Onteru SK, Singh D: **Genomic technologies: a way forward for learning climate resilience through cellular responses to heat stress**. *Climate Resilient Livestock & Production System* 2012, **1**(5).

10. Kang HM, Sul JH, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E: **Variance component model to account for sample structure in genome-wide association studies**. *Nature genetics* 2010, **42**(4):348.

11. Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y: **The channel catfish genome sequence provides insights into the evolution of scale**

formation in teleosts. *Nature communications* 2016, **7**:11757.

12. Stanke M, Morgenstern B: **AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints**. *Nucleic acids research* 2005, **33**(suppl_2):W465-W467.

13. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic acids research* 2005, **33**(suppl_1):D501-D504.

14. Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y, Zhang J, Feng J, Kaltenboeck L, Kucuktas H: **Development of the catfish 250K SNP array for genome-wide association studies**. *BMC research notes* 2014, **7**(1):135.

15. Zhou X, Stephens M: **Genome-wide efficient mixed-model analysis for association studies**. *Nature genetics* 2012, **44**(7):821.

16. Luo C, Qu H, Ma J, Wang J, Li C, Yang C, Hu X, Li N, Shu D: **Genome-wide association study of antibody response to Newcastle disease virus in chicken**. *BMC genetics* 2013, **14**(1):42.

17. LaFrentz BR, Shoemaker CA, Booth NJ, Peterson BC, Ourth DD: **Spleen index and mannose-binding lectin levels in four channel catfish families exhibiting different susceptibilities to *Flavobacterium columnare* and *Edwardsiella ictaluri***. *Journal of Aquatic Animal Health* 2012, **24**(3):141-147.

18. Ledur M, Navarro N, Pérez-Enciso M: **Large-scale SNP genotyping in crosses between outbred lines: how useful is it?** *Heredity* 2010, **105**(2):173.

19. Logan CA, Somero GN: **Transcriptional responses to thermal acclimation in the eurythermal fish *Gillichthys mirabilis* (Cooper 1864)**. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2010, **299**(3):R843-R852.

20. Logan CA, Somero GN: **Effects of thermal acclimation on transcriptional responses to acute heat stress in the eurythermal fish *Gillichthys mirabilis* (Cooper)**. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* 2011, **300**(6):R1373-R1383.

21. Quinn NL, McGowan CR, Cooper GA, Koop BF, Davidson WS: **Identification of genes associated with heat tolerance in Arctic charr exposed to acute thermal stress**. *Physiological Genomics* 2011, **43**(11):685-696.

22. Castilho PC, Buckley BA, Somero G, Block BA: **Heterologous hybridization to a complementary DNA microarray reveals the effect of thermal acclimation in the endothermic bluefin tuna (*Thunnus orientalis*)**. *Molecular ecology* 2009,

18(10):2092-2102.

23. Buckley BA, Gracey AY, Somero GN: **The cellular response to heat stress in the goby *Gillichthys mirabilis*: a cDNA microarray and protein-level analysis**. *Journal of Experimental Biology* 2006, **209**(14):2660-2677.

24. Luo GR, Le WD: **Collective roles of molecular chaperones in protein degradation pathways associated with neurodegenerative diseases**. *Current pharmaceutical biotechnology* 2010, **11**(2):180-187.

25. Fuquay J: **Heat Stress as it Affects Animal Production 1**. *Journal of animal science* 1981, **52**(1):164-174.

26. Matsuki S, Iuchi Y, Ikeda Y, Sasagawa I, Tomita Y, Fujii J: **Suppression of cytochrome c release and apoptosis in testes with heat stress by minocycline**. *Biochemical and biophysical research communications* 2003, **312**(3):843-849.

Table 1 The significantly associated SNPs on linkage group 14 and 16. R^2 indicates the ratio of phenotypic variation of SNPs. MAF refers to minor allele frequency.

Linkage group	SNP ID	Position	Allele	-$\log_{10}$(P-value)	MAF	R^2	Nearest gene
14	AX-85318076	14 002 635	C/A	5.69	0.25	0.12	LIFR
16	AX-86115098	15 552 403	G/A	5.44	0.32	0.12	COL12A1
16	AX-85265807	18 651 314	T/C	5.36	0.15	0.12	PLCB1

Table 2 Information of genes associated with heat stress

Linkage group	SNP ID	Position (bp)	$-\log_{10}$ (P-value)	Gene name	Gene position	Biological process
14	AX-85318076	14,002,635	5.69	TNF receptor-associated factor 2	13 781 935-13 794 008	Protein degradation
				F-box/WD repeat-containing protein 5	13 798 581-13 805 052	Protein degradation
				Anaphase-promoting complex subunit 2	13 855 865-13 864 890	Protein degradation
				Solute carrier family 25 member 46	14 419 301-14 432 323	Molecule transport
				U4/U6 small nuclear ribonucleoprotein Prp4	14 505 527-14 510 937	Protein biosynthesis
				DnaJ homolog subfamily C member 25	14 564 010-14 567 877	Protein folding
16	AX-86115098	15,552,403	5.44	Collagen alpha-1(XII) chain	15 534 822-15 554 851	Cytoskeletal reorganization
				Cytochrome c oxidase subunit 7A2	15 561 184-15 565 412	Energy metabolism
				E3 ubiquitin-protein ligase UBR1	15 808 218-15 829 181	Protein degradation
16	AX-85265807	18,651,314	5.36	Chloride intracellular channel protein 5	18 145 365-18 150 228	Iron transport
				Kelch-like protein 29	18 398 355-18 411 722	Protein degradation
				1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-1	18 637 307-18 647 131	Energy metabolism
				1-phosphatidylinositol 4,5-bisphosphate phosphodiesterase beta-4	18 683 546-18 702 211	Energy metabolism
				Heterogeneous nuclear ribonucleoprotein Q	18 996 404-19 002 665	Protein biosynthesis

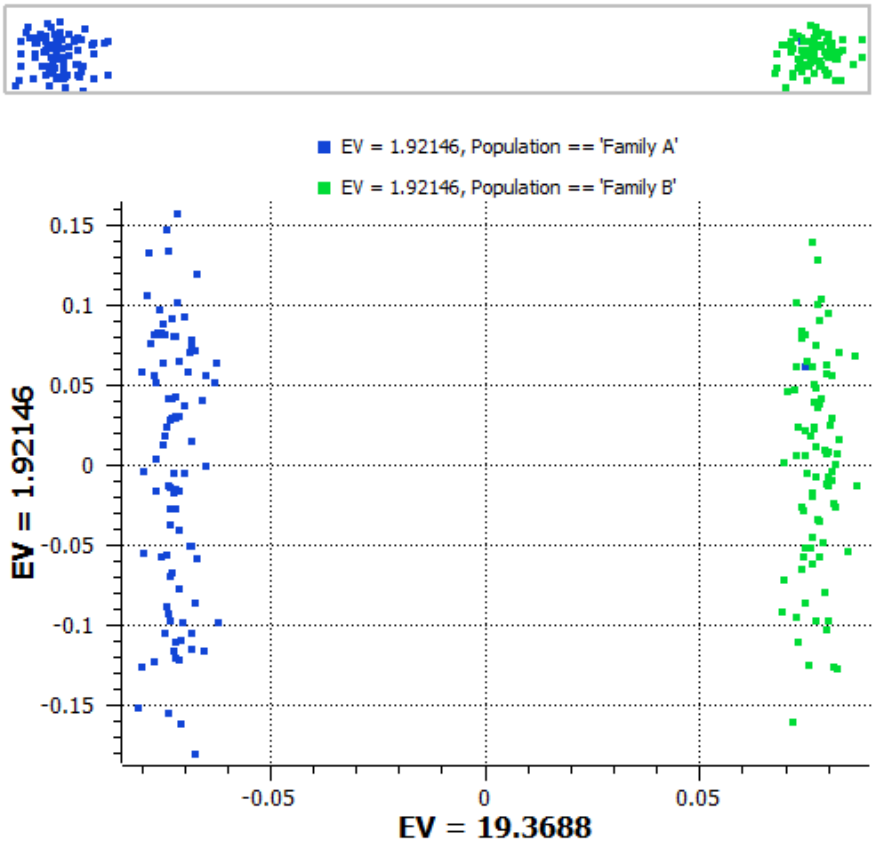

Figure 1. Sample structure identified by PCA with the first two principal components. The coordinates are the first two principle component scores.

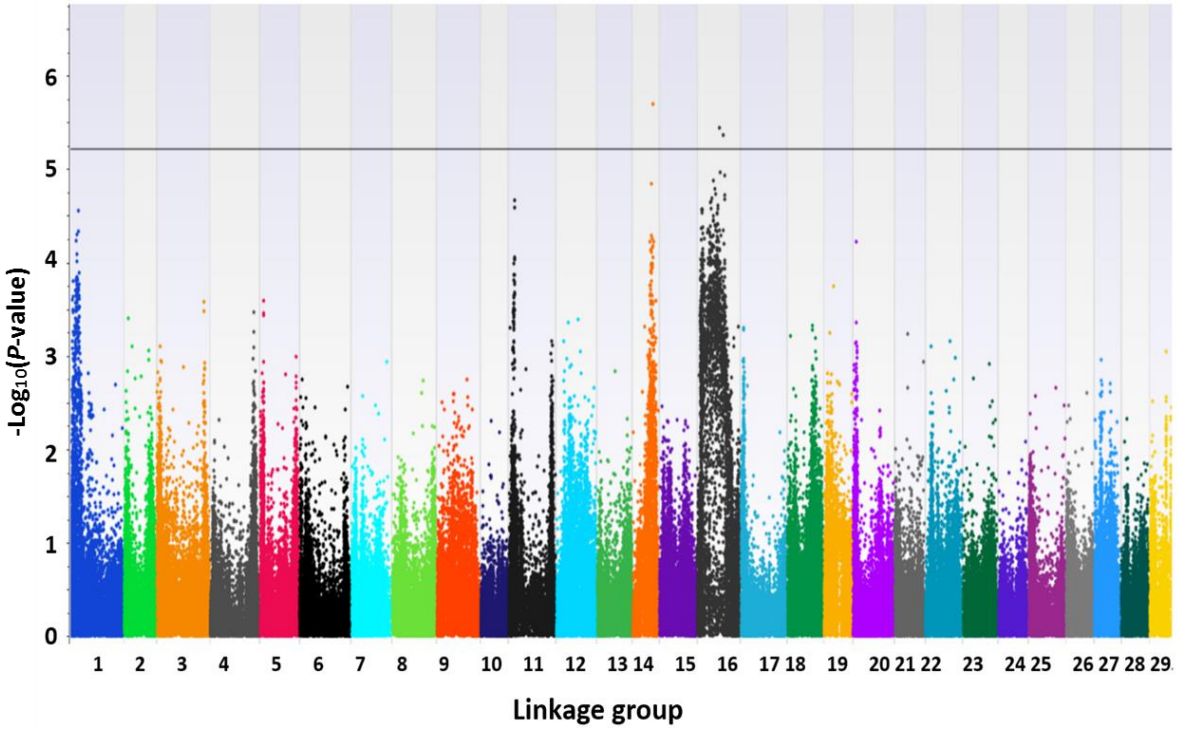

Figure 2. A Manhattan plot of genome-wide association analysis for heat stress. GWAS identified three loci associated with heat stress in catfish. The horizontal line indicates the genome-wide significant threshold: $-\log_{10}$ (P-value) = 5.209.

Chapter 3. Analysis disease resistance against enteric septicemia of catfish (ESC) using BSR-Seq analysis

Abstract

Identification of genetic markers to genes of ESC (Enteric septicemia of catfish) resistance is a major task of catfish research and breeding program. Bulked segregated RNA-Seq (BSR-Seq) approach appears to be a cost-effective solution to mapping genes and QTLs for many economic traits in a rapid and efficient manner. Significantly differentially expressed SNPs and global gene expression patterns in highly relevant tissues would lead us identify causal loci underling the traits of interest. In this study, Channel catfish liver and intestine samples were parallelly analyzed to characterize transcriptomic profiles and identify significantly differentially expressed SNPs following *E. ictaluri* infection using state-of-art BSR-Seq. Transcriptomic profiling analysis revealed divergent cellular responses between liver and intestine after infection, demonstrating they could collaborate closely while keep hemostasis using own specific strategies and potential unique mechanisms after infection. Immune/inflammatory related processes or pathways were only enriched from liver and more intensive immune and defense responses were observed in resistant fish than susceptible fish, indicating liver exerts crucial roles in ESC resistance. The methodology of Euclidean distance (ED) allowed to identify five significant SNPs with large ED values on catfish genome. Three significant SNPs were from liver and physically linked on LG1. The other two were from intestine and located on LG12 and LG26, respectively. Collectively, these significant SNPs suggest three associated QTLs with ESC resistance, which are successfully verified our

previous GWAS studies. Eleven genes were found to be differentially regulated between resistant fish and susceptible fish within the three QTL regions, indicating their important involvements in disease resistance. Of particular interest is the Apo-14 kDa gene displayed both differentially up-regulation and significantly allelic segregation of SNP between resistant fish and susceptible fish. Catfish Apo-14 kDa gene is homologue of Apoa-II, which was recognized as an antimicrobial protein in mammals and teleost fish. With this respect, Apo-14 kDa might also have antimicrobial activity in catfish and could be a promising candidate for ESC resistance. The present study facilitates the knowledge of further marker-assisted selection/genome-based selection and gene knock-out techniques for ESC resistance in catfish.

Introduction

Enteric septicemia of catfish (ESC), caused by a Gram-negative bacterium *Edwardsiella ictaluri*, is the most severe disease problem for the catfish industry. The bacterium infects both juvenile and adult food-size fish, and therefore, can cause the whole loss of the operations for catfish producers. Various disease control approaches have been explored for the control of ESC disease, but none of the methods are completely effective. Vaccines for ESC disease was developed and commercially available [1], but the vaccine is effective only for a selected fraction of isolates of the pathogen. An antibiotic, commercially marketed as Aquaflor (effective component florfenicol), was approved by FDA in 2005 for the control of ESC. It is effective, but use of large quantities of antibiotics in aquatic environment has adverse environmental impact, increasing human health risks.

Channel catfish (*Ictalurus punctatus*) is the major cultivated species of the US catfish

industry. However, in recent years, the hybrid catfish (channel catfish female x blue catfish male) has been increasingly cultivated in the catfish industry, and now the accounts for over 60% of the US catfish industry [2]. In terms of ESC resistance, channel catfish is generally susceptible to ESC infection, while blue catfish (*I. furcatus*) is generally highly resistant [3-5]. However, within channel catfish populations or families, large variations do also exist in disease resistance against ESC disease. The genetic variations both between species and within species make genetic improvement of disease resistance an effective approach for ESC disease control. Therefore, understanding the molecular mechanisms underlying ESC disease resistance is of great interest to enhance resistance using genetic approaches.

Several strategies have been used to study molecular mechanisms and identify genes responsible for ESC disease resistance in catfish. These included both genomic and transcriptomic levels of analyses. Three genome-wide association studies (GWAS) were conducted to identify QTL responsible for ESC disease, using various populations including intraspecific channel catfish families [6], the second generation of interspecific backcross hybrid catfish [7] and the fourth generation of interspecific backcross hybrid catfish [8]. In these studies, both interspecific and intraspecific QTL were identified. For instance, major disease resistance QTL were identified on LG1 and the resistance alleles were of the blue catfish origin [7, 8]. Similarly, major QTLs were also identified among channel catfish populations that were located on LG1 and LG26 [6]. These analyses were highly effective for the determination of genomic regions that include major ESC disease resistance QTL. However, fine mapping of these QTL was limited by the low recombination rate in the mapping families. Genetically unrelated populations should be more effective, but such populations often are

difficult to obtain for aquaculture species. In addition, analysis of a large number of samples using high-density SNP arrays can be very expensive with aquaculture species.

Transcriptome level analyses were conducted to determine differentially expressed genes (DEGs) after infection. A microarray study was conducted to identify DEGs early after infection [9]. The significantly up-regulated genes were those of complement-related components, pathogen recognition receptors (PRRs) and CC chemokines [9]. Later on, a RNA-Seq analysis allowed the identification of 1,633 DEGs at 3 h, 24 h, and 3 days following *E. ictaluri* infection [10]. Of the DEGs, several actin-related genes were notably found after infection, such as genes encoding ARP2/3, ezrin, filamin, Rho-GTPase, CDC42SE2, integrins and gelsolin-like. Dramatic dysregulation of components of the apical junction complex (AJC) was detected and represented by the claudin genes, desmocollin 2, desmoplakin, epithelial cadherin precursor and magi3 [10]. At the same time, a BSR-Seq analysis using liver of F2 backcross catfish progenies observed dramatically induced immune-related DEGs between resistance and susceptible fish, such as apolipoproteins, complement components, fibrinogen alpha and gamma, and MHC class genes [11]. These DEGs coupled with disease resistance QTL locations, provided insights into the potential mechanisms of disease resistance, but the analysis was conducted in different populations.

Combining the genomic and transcriptomic analyses should be more powerful. One of such analysis is the bulked segregant RNA-Seq (BSR-Seq). BSR-Seq is an integrated solution of bulk segregant analysis (BSA) and RNA-Seq technology. BSA is capable of identifying genetic markers related with any specific gene or genomic regions using two bulks of pooled DNA samples with contrasting phenotypes [12]. This approach appeared to be flexible as

different types of segregating populations can be used to generate extreme phenotypes [13]. Advances in NGS allowed effective discovery of massive molecular markers such as single-nucleotide polymorphism **(**SNPs), which promoted BSA in broader applications of NGS-based studies [14]. Therefore, coupling with RNA-Seq, BSR-Seq appeared to be a powerful tool to map genes and QTLs underling economic traits in a rapid and efficient manner. Through multiplying the advantages from both the BSA and RNA-Seq, theoretically, BSR-Seq has the full capability to identify DEGs between bulks, and also the ability to map significant SNPs to genes associated with the traits [11]. Because of such advantages, BSR-Seq has been widely applied in both plants [13, 15-17] and animals [11, 18].

BSR-Seq analysis on animals was first employed on ESC resistance of catfish using F2 backcross catfish progenies, which allowed identification of numerous DEGs, as well as putative genomic regions involve in ESC disease resistance [11]. However, without a well assembled reference genome sequence, Wang et al. (2013) could not map the significantly differentially expressed SNPs to precise genomic locations, prohibiting analysis of QTL using BSR-Seq [11]. With the accomplishments of a high-quality catfish genome assembly [19] and high-density genetic linkage maps [20, 21], our ability to map the significantly differentially expressed SNPs is drastically enhanced. The primary objective of the present study was to determine ESC resistance-associated SNPs within channel catfish using BSR-Seq methodology. The secondary objective was to determine if similar results can be obtained from BSR-Seq and GWAS analysis, and if so, what positional and expressional genes could be responsible for ESC resistance.

Transcriptomic profiling of the liver and intestine has usually been utilized for measuring

gene responses to diseases and revealing the immune system in aquaculture species, such as in catfish [9-11], Atlantic salmon [22-24], trout [25], tilapia [26] [27] and carp [28, 29]. Catfish intestine tract is the primary infection site of *E. ictaluri*, which plays direct and critical function in cytoskeletal dynamics and junctional regulation in pathogen entry and subsequent inflammatory responses [10]. Liver has been shown to be one of the most important classical immune organs in catfish as well, which likely plays an important function in pathogen defense as well as inflammatory signaling [9, 11]. In particular, liver contains essential and specialized immune tolerance mechanisms to degrade and to clean antigens and endotoxins from the gut microbiota [30]. For example, there are several specialized immune or immune-related cells (dendritic cells, Kupffer cells, hepatic stellate cells, natural killer cells and T regulatory cells) enriched in the liver [31]. Here in this study, we are using both liver and intestine tissues for BSR-Seq analysis.

Materials and methods

ESC bacteria challenge and sample collection

All procedure involving the handling and treatment of fish used during this study followed the protocols approved by Auburn University Institution Animal Care and Use Committee (AU-IACUC). Three families of channel catfish were used to conduct challenge, with treatment group containing 900 fish (300/family) and control group containing 300 fish (100/family). Cultures of a bacterial isolate MS-S97-773 of *E. ictaluri* was used for challenge experiment. Bacteria was cultured from a single colony, re-isolated from a symptomatic fish, and confirmed biochemically before being inoculated into brain heart infusion (BHI) medium and incubated in a shaker incubator at 28 °C overnight. The bacterial concentration was determined using

colony forming unit (CFU) per ml by plating 10 µl of 10-fold serial dilutions onto BHI agar plates. For each family, the parallel treatment of bacteria challenge was conduct as follows: 300 progenies were placed in a static 300-L tank for bacteria challenge. A 300-ml bacteria culture containing 1.1 x 10^9 *E. ictaluri* cells/ml was added to the tank, resulting in a final concentration of 1.1 x 10^6 cells/ml. The treatment fish were challenged by immersion bath for two hours, and then provided continuous water flow-through during the challenge experiment. The control fish were treated with identical procedures except that they were exposed to sterile BHI medium.

The strategy of BSA was employed in bulks design and sample collection. Each sample was collected independently. Earliest mortalities of 12% dying fish (36 fish per family) with classical ESC clinic signs were collected as susceptible bulk from the three families. After two weeks of the challenge, the survival fish were regarded as resistant fish. Samples from 36 randomly picked survivors (12% of the treatment fish) were collected as the resistant bulk in each of the three families. Besides, 12 random individuals from each control group of the three families were obtained before bacteria challenge and later combined as control bulk (also 36 samples in total). Tissues of liver, and intestine were collected for each sample, and stored at −80 °C until RNA extraction.

RNA isolation and sequencing

RNA-seq was carried out by using liver and intestine tissues collected from the six phenotypic bulks (three susceptible bulks and three resistant bulks) and one control bulks. Equal amount of tissues from each fish sample was used for RNA extraction independently. RNA was extracted using RNeasy Mini kit (Qiagen, USA) following the manufacturer's

protocol. The concentration and integrity of each RNA sample were measured using a NanoDrop ND-1000 UV-VIS spectrophotometer (NanoDrop Technologies, USA). The RNA samples belonging to the same bulk were diluted to the same concentration and then pooled together prior to cDNA library construction. Taken together, 14 RNA samples were prepared for library construction and subsequent sequencing.

The 14 libraries were constructed and sequenced on a Hiseq 2500 instrument with 125 bp paired end (PE) according to manufacturer's instructions (Illumina, San Diego, CA, United States). Briefly, Poly-A RNA containing mRNA was enriched using poly-T oligo-attached magnetic beads and fragmented. Second-strand cDNA was synthesized using reverse-transcriptase (Invitrogen) kit and random hexamer primers, then purified, end-repaired, poly-A tailed, and adaptor ligased. The cDNA pools were loaded to and paired-end sequenced on a Illumina Hiseq 2500 sequencer (Illumina, United States) at HudsonAlpha Genomic Services Lab (GSL).

Reads mapping, and expressed transcripts assembly

Raw reads were first assessed for their quality using the FastQC (Version 0.11.5) [31]. Clean reads were obtained by discarding adaptor sequences, ambiguous nucleotides ('N' in the end of reads), extreme short reads (<36 bp), and low-quality reads (phred quality score of less than 20) via Trimmomatic (version 0.33) [32]. Then, the clean reads for each sample were aligned to the catfish reference genome [18] through Tophat2 [33], with the minimum alignment of 90% and the maximum of two base mismatches. The alignments of each sample were stored as a single BAM file followed by transcripts assembly using Cufflinks [34]. In addition, Cuffmerge was applied to perform a reference annotation-based transcripts (RABT)

assembly.

Differentially expressed genes identification

The Identification of DGEs was conducted by Cuffdiff [34], which treated the phenotypical bulks (susceptible/resistant) from the three families as bulked replicates to control the variation of reads counts for each gene across the families, as well as calculate the significance of detected changes in expression by using these estimated variations, thereby improving the abundance estimates. Briefly, for each tissue, BAM files from three susceptible bulks were treated as three replicates and combined to form the S bulk, similarly, the R bulk was generated from the resistant bulks, and C bulk was defined as control bulk. Together, DEGs were identified between S vs. C bulks, R vs C bulks, and R vs. S bulks, respectively. During the gene abundance analysis, total mapped reads were counted for each gene and then normalized to produce fragments per kilobase of exon model per million mapped reads (FPKM). The statistical significance of differential expression was determined using multiple testing combined with false discovery rate (FDR). In this study, only genes with more than two-fold change in relative expression and p-value <0.01 (FDR <0.05) were classified as DEGs.

GO terms enrichment analysis using DAVID and ClueGo

Functional annotation analyses of the DEGs between resistance fish and susceptible fish were firstly performed using the Functional Annotation tool within the DAVID Bioinformatics Database (http://david.abcc.ncifcrf.gov/home.jsp). The default annotation categories and the threshold of 2 counts were chosen to minimize exclusion of potentially interesting functional terms. The terms or pathways with a modified Fisher-exact P-value (EASE score) < 0.01 were considered as significantly enriched. To gain more comprehensive insights about the DEGs

enriched patterns that can integrate interesting DEGs in this study, we also performed network-based analysis of GO and KEGG pathway enrichment using Cytoscape version 3.5.1 (http://www.cytoscape.org/) [42] with the ClueGo version 2.3.3 plugin (http://www.ici.upmc.fr/cluego) [43]. For this study, the 'GO term fusion' strategy was adopted and 'show only significant GO terms (Bonferroni corrected P-value < 0.01)' were set to consider only significant enriched groups.

Identification of SNPs and significant SNPs

Initially, all the putative SNP were called by using SAMtools 1.5 [35] and Popoolation 2 [36]. In this step, BAM files from the three susceptible and three resistant bulks generated from liver were merged to create a single S bulk and R bulk, respectively. Such merged data can produce greater depth of sequencing data, providing more reliable results in discovery of variants. BAM files for S bulk and R bulk of intestine were generated by the same way. After duplicate removal using SAMtools' rmdup command, the new generated BAM files belonging to same tissue (one set of S bulk and R bulk for liver, another set S bulk and R bulk for intestine) were combined accordingly through the mpileup function to form a single mpileup file. Thereafter, A java module (mpleup2sync.jar) in Popoolation2 was employed to detect putative SNPs sites between S bulk and R bulk by using the mpileup file with the following parameters, which mainly based on sequencing coverage and reads quality: (1) the sequencing depth was more than 6X, and (2) the quality value was larger than 30. Secondly, a perl script (snp-frequency-diff.pl) in PoPoolation2 was employed to call the genotype at every putative SNPs, and to compare allele frequency differences between the phenotypic bulks. Thirdly, the significant SNPs were identified with showing only two allele variants, as well as allele

differences were significantly different between the two bulks after two-tailed Fisher's Exact test. The threshold was set as FDR p-value ≤ 0.05. However, the statistical significance of SNPs is compromised by the expression levels. The highly expressed genes had much lower p-values. Therefore, the Fishers exact test may be used to identify significant changes in allele frequency.

Euclidean distance calculation

The methodology of Euclidean distance (ED) was employed to further measure allele segregation of significant SNPs more directly. ED is metric that has advantages such as without the requirement of parental strain information and resistant to the sequencing noise of RNA-Seq. Using the equation:

$$\boldsymbol{ED = \sqrt{(A_R - A_S)^2 + (C_R - C_S)^2 + (G_R - G_S)^2 + (T_R - T_S)^2}}$$

Where each letter (A, T, C, G) represents the frequency of its corresponding DNA nucleotide. In order to reduce noise generated from small variations when estimating the allelic frequency, the ED value was raised to a power 5 mode. The data were recognized as the representatives of correlation effects, and plotted using qqman [5].

Identification of regions associated with ESC resistance in channel catfish

In this study, $ED^5 > 1$ was selected as threshold to identify significantly SNPs associated with ESC resistance. Genomic regions within ±1 Mb of ED value peaks were regarded as candidate regions. Genes within candidate regions were predicted from the catfish genome sequences [18] by AUGUSTUS [37] and FGENESH [38], and annotated by BLAST analysis and non-redundant (nr) protein database [39]. Gene function and pathways were collected form Uniprot (http://www.uniprot.org) and GeneCards (http://www.genecards.org) databases.

Results

RNA-Seq reads mapping and transcriptome assembly

RNA-Seq of 14 bulked libraries was conducted using Illumina paired-end sequencing platform using specific barcodes to trace and analyze the reads separately. The reads number of Illumina sequencing and mapping information of these 14 samples were summarized in Table 1. As a whole, 675 million raw reads were generated with similar reads numbers from liver and intestine. After quality trimming, over 554 million clean reads were retained for transcript mapping. The average transcript mapping rate was about 83% and 78% for the liver and intestine bulks, respectively. The mapped reads accounted for 12,748 to 17,695 genes across each bulk from the liver and intestine samples, respectively.

DEGs in liver and intestine after infection

Transcriptomic profiles were built by comparative analysis of: 1) between susceptible bulk and control; 2) between resistant bulk and control; 3) between resistant bulk and susceptible bulk; for both the liver and intestine tissues. Differentially expressed genes (DEGs) were determined by their expression levels based on FPKM (fragments per kilobase of exon model per million mapped reads). The DEGs from the six comparisons are summarized in Table 2. For the liver, 3,092, 1,652, and 607 genes were identified from comparisons of S vs. C, R vs. C and R vs. S, respectively. For the intestine, 2,241, 917, and 678 DEGs were identified for the S vs. C, R vs. C and R vs. S comparisons, respectively. The vast majority of differentially expressed genes were between fold change 2-5. However, many genes were drastically expressed after infection in different comparisons. For instance, of the genes significantly up-

regulated in liver after infection, 36 genes were expressed over 50-fold in susceptible fish than in control fish; 25 genes were regulated more than 50-fold in resistant fish than in control fish; 8 were expressed 10-50 times more in resistant fish than in susceptible fish. Of the genes dramatically up-regulated in intestine, 6 genes were expressed over 50-fold in susceptible fish than in control fish; 26 were expressed 10-50 times higher in resistant fish than in control fish; 4 genes were expressed 10-50 times more in resistant fish than in susceptible fish.

Across these six comparisons (Table 2), the number of DEGs in liver was much greater than in intestine after infection. By comparing with control fish, both the susceptible and resistant fish contained drastically more number of DEGs in liver (3092 and 1652) than that in intestine (2241 and 917). Moreover, the number of up-regulated DEGs was much greater in liver as well. For example, among these DEGs in resistant fish compared with control fish, the number of up-regulated DEGs (1001) was nearly two times greater than down-regulated genes (651) in liver samples, while in intestine samples, the number of up-regulated DEGs (320) was much less than the number of down-regulated DEGs (597). Importantly, DEGs between phenotypic bulks (R vs. S) were mostly concerned and emphasized. In this case, the number pattern of up- and down-regulated DEGs between liver and intestine samples were opposite: 345 up-regulated DEGs vs. 262 down-regulated DEGs were identified in liver, but 277 up-regulated DEGs vs. 401 down-regulated DEGs were found in intestine.

Enrichment analysis of DEG between resistant and susceptible fish in liver and intestine

DEGs between phenotypic bulks (R vs. S) were conducted for the enrichment analyses, which is a direct representation of the molecular responses correlated with ESC resistance. Firstly, DEGs from liver and intestine samples were used as two independent inputs for the

functional annotation analysis with DAVID, which revealed several different enriched GO terms between the two tissues (Figure 1A & 1B). The number of DEGs under various GO terms was indicated as well. In liver, nine GO terms were significantly enriched, including oxidation-reduction process, lipid metabolic process, inflammatory response, response to lipopolysaccharide, glycolytic process, immune system process, fatty acid metabolic process, and regulation of cell death (Figure 1A, Table 3). Immune system process and inflammatory response are critical steps to adaption to bacteria infection and maintain immune hemostasis. With respect to immune system process, six DEGs fall in this GO terms with four genes were up-regulated in resistant fish than in susceptible fish. Thirteen DEGs fall in the inflammatory response GO term, with 10 genes induced. In intestine, seven enriched GO terms were enriched that contained oxidation-reduction process, proteolysis, negative regulation of transcription, regulation of cell growth, response to bacterium, cholesterol biosynthetic process, and protein polymerization (Figure 1B, Table 4). Response to bacterium represents any state or activity of a cell or an organism as a result of a stimulus from a bacterium. Under this GO term, eight DEGs were detected with 6 genes were up-regulated in susceptible fish than in resistant fish. The common enriched GO term between liver and intestine was oxidation-reduction process, which contained much more number of DEGs than the other GO terms, indicating its vital role in the host for ESC resistance/susceptibility. In detail, 36 DEGs from liver were identified under the oxidation-reduction process, 16 and 20 of these genes were up- or down-regulated in resistant fish than susceptible fish, respectively. Similar number of DEGs (40) from intestine was under this GO term, with 17 were expressed higher and 23 were expressed lower in resistant fish than in susceptible fish. According to above enriched GO terms, the transcriptome

changes between liver and intestine after infection showed large differences but minor similarities (Figure 1A & 1B). The immune related terms were only enriched from liver DEGs, which could elicit the immune importance of catfish liver after ESC infection.

Furthermore, the resistant fish up- and down-regulated DEGs were then treated as independent inputs into ClueGO in Cytoscape for integrated functional analysis. The results clearly demonstrated divergent transcriptomic differences between liver and intestine after ESC infection (Figure 2A & 2B). As shown in Figure 2A, 13 functional enriched groups were revealed from liver DEGs. Among these groups, 11 of them were enriched by both up- and down-regulated DEGs with the up-regulated DEGs taking most of the percentages. The other two groups, purine metabolic process and response to cAMP, were revealed specifically from the up- and down-regulated DEGs, respectively (Figure 2A). Four of these 13 functional groups are involved in immune systems, including response to cytokine, defense response, regulation of immune system process, and PPAR signaling pathway. The resistant fish up-regulated DEGs were dominant in these four immune groups, indicating the basic immune and defense processes at steady state as well as intensive defense response in catfish liver after ESC infection. For intestine DEGs, seven functional groups were enriched by both up- and down-regulated DGEs (Figure 2B). No immune-related functional groups were identified from intestine DEGs. However, groups of response to wounding healing, protein degradation and synthesis (proteolysis, endopeptidase inhibitor activity), and cell growth process (insulin-like growth factor binding) were specifically revealed from intestine DEGs. The majority of representatives to each group were from down-regulated DEGs in resistant fish than susceptible fish, representing massively molecular responses and gene regulation of intestine toward

susceptible fish after ESC infection.

Significant SNPs and QTL associated with ESC disease resistance

The power of BSR-Seq for the analysis of QTL is offered by differential allelic representation of SNPs in the contrast phenotypic bulks. In order to identify potential SNPs associated with QTL, the first step was to identify SNPs detected in the transcriptome analysis. Overall, 145,377 and 206,328 SNPs were identified in the phenotypic bulks from liver and intestine, respectively. Of these, a total of 10,790 (7.4%) and 12,986 (6.3%) SNPs were found to be differentially represented between the resistant and susceptible bulks at FDR of 0.05 in the liver and intestine, respectively (Table 5).

To further determine if the significant SNPs were associated with QTL, Euclidean distance (ED) analysis was conducted to calculate frequency distance. Since ED values ranged from 0 to $\sqrt{2}$, we used its 5th power (ED^5) to sufficiently increase the effect of large ED measurements from associated locus and reduce the effects of low ED measurements from small variations in allelic frequency estimation. Using significant SNPs from the liver, only one region containing ED^5>1, on LG1. Three significant SNPs that are physically linked with large ED^5: ED^5 of 2.8 at position 7,491,537; ED^5 of 2.1 at position 7,491,492; and ED^5 of 1.4 at position 7,468,885 (Figure 3, Table 6), suggesting an associated QTL in this region.

Using significant SNPs from intestine, two regions containing SNPs greater than 1 were identified, one on LG12 and the other on LG26 (Figure 2). ED^5 on LG12 was 1.3 at position 24,564,284; and ED^5 on LG26 was 1.1 at position 5,146,464 (Figure 4, Table 6), suggesting an associated QTL in LG12 and LG26.

Comparative analysis of QTL identified from BSR-Seq and GWAS

Comparative plots of SNPs along LG1 from BSR-Seq (this study) and GWAS analysis [6] are displayed on Figure 5. The SNPs with $ED^5 > 1$ were distributed between position 7,468,885-7,491,537. The two most significant SNPs (LG1: 7,491,492 and 7,491,537) were located within apolipoprotein C-I (Apoc1) gene, and the third SNP (LG1:7,468,885) was located in the gene encoding 14KDa apolipoprotein (Apo-14kDa) (Figure 5A, Table 6). In comparison, previous GWAS identified a major genomic region of 3.7 Mb (from 5,948,310 to 9,638,510 bp) on LG1 that is significantly associated with ESC resistance in channel catfish (Figure 5B), which is overlapped with the region identified by BSR-Seq here.

On LG12, ED analysis highlighted an associated region within 24-26 Mb (Figure 6A). The SNP (LG12: 24,564,284) with largest ED value is located close to the of RNA polymerase III subunit H gene (POLR3H) (Table 6). In addition to the SNP with ED^5 larger than 1, five additional SNPs had ED^5 values apparently larger than their surrounding SNPs at position 24,044,560-24,606,902. Taken together, these SNPs with large ED^5 values suggested the presence of an associated QTL in this region on LG12. The GWAS study in channel catfish [6] didn't identify any QTL in this region on LG12. However, one suggestive QTL was identified around position 24 Mb on LG 12 in the forth generation of hybrid catfish [8] (Figure 6B).

On LG26, we observed one association region around 5 Mb (Figure 7A), which carries one significant SNP with large ED value in the gene of NADPH oxidase activator 1 (NOXA1). In addition to this SNP with $ED^5>1$, several additional SNPs with smaller ED^5 were also identified in the same gene. With GWAS analysis, one significant QTL was identified in LG26, although the most significantly associated region was found at position between1,639,485 to

1,666,617 bp, another suggestively associated region was found between 4,765,338 to 5,379,481 bp [6] (Figure 7B).

Differentially expressed genes within the associated QTL regions

To determine the intersection of genes that are both located in the associated QTL regions and also differentially expressed between the resistant and susceptible bulks, we first determined all the genes within 2 Mb (1 Mb upstream and down-stream of the most significant SNP). A total of 58 genes were located within the QTL region (6.5-8.5 Mb) on LG1. Of the 58 genes, 46 genes were expressed in the liver, and 5 of the 46 genes were differentially expressed between the resistance and susceptible bulks (Table 8). Of these five genes, 14KDa apolipoprotein (Apo-14kDa) was expressed significantly higher (2.5-fold) in resistant fish than in susceptible fish, while the other four genes, ictacalcin (40.0-fold), GTP-binding protein 10 (GTPBP10) (3.4-fold), apolipoprotein Eb (Apo-Eb) (2.6-fold), and ephrin A1 (EFNA1) (2.4-fold), were expressed higher in susceptible fish than in resistant fish (Table 7).

On LG 12, 65 genes were located in the QTL region (23.1-25.1 Mb) (Table 9). The expression analysis detected 43 expressed genes from intestine with 3 DEGs were observed, including XK-related protein 8 (XKR8), ATP-dependent RNA helicase DDX39A (DDX39A), and Traf2 and NCK-interacting protein kinase (TNIK). All these three genes were dramatically up-regulated in susceptible fish than in resistant fish at 2.0-fold, 2.4-fold, and 2.5-fold, respectively (Table 7).

On LG26, we identified 76 genes in that QTL region (4.1-6.1 Mb), among which 60 expressed genes were observed from intestine after infection with 3 genes were differentially expressed (Table 10). The three DEGs are C-C motif chemokine 19 (CCL19), serine protease

inhibitor 2.1 (SPI2.1), and Dnaj homolog subfamily B member 5 (DNAJB5). Of these three genes, CCL19 and DNAJB5 were expressed higher in resistant fish than in susceptible fish at 3.9-fold and 1.9-fold, respectively. The gene coding SPI2.1 protein was expressed higher in susceptible fish than in resistant fish with 2.2-fold (Table 7).

Discussion

ESC is the most destructive disease to catfish industry that causing substantial economic losses each year. In general, channel catfish is very susceptible to ESC that often results in heavy mortalities as early as four days after onset of infection [9]. The stress-related reductions in immune function after ESC infection have been largely attributed to the durative increases of infectivity and mortality of channel catfish [32]. However, some resistant fish do exist from different populations. Therefore, identification and isolation of major QTL and causal genes responsible for ESC resistance would be one of the most straightforward and useful ways for improving the ESC resistance in catfish industry. BSR-Seq provides an integrated method to both discover genetic variants and DEGs between the phenotypical bulks. Thereby, genome-wide information about positional candidates and potentially functional genes associated with the trait of interest can be obtained in a reliable, and cost-effective way [33]. In this study, pooling strategy of BSR-Seq was applied to amplify the genetic signals between phenotypic extremes but reduce variations from non-target traits within and between bulks, allowing us to focus on disease susceptibility/resistance for ESC in catfish. The BSR-Seq datasets not only allowed to identify hundreds of genes differentially expressed between ESC resistant fish and susceptible fish using liver and intestine tissues, but also localized ESC disease resistance-

associated regions on the catfish genome.

Considering the cost-effective features of BSR-Seq, it could be a good alternative for QTL analysis without genotyping a large number of samples, such as using expensive SNP arrays. Three main advantages are offered by BSR-Seq based mapping in comparison with GWAS: firstly, RNA-Seq allows to direct comparison of gene expression levels between phenotypical bulks, to understand transcription mechanisms, and to identify potential candidate genes whose expression levels are influenced by mutations in the coding regions or caused by regulatory mutations [18]. Secondly, transcriptomic sequencing allows to visualize the annotated or unannotated isoforms or differential splicing sites. Finally, RNA-Seq approaches experience a dramatically decreased cost compared with GWAS. A number of powerful bioinformatic tools are well available for RNA-Seq data sets and been extensively used. However, we still recognize the limitations of BSR-Seq. For example, it is unable to examine the polymorphisms from intron or noncoding regions. With this respect, further targeted genomic sequencing would be a complementary way to identify causal loci in noncoding regions surrounding the expression candidates. It is unable to find the associated loci if the transcripts of interest are missing from the RNA library, or expressed at low levels at the collection time points. In this case, the best solution is to collect RNA sample at the highly correlated tissues and at timepoints of the first emergency of the phenotype that are interested, which help increased the possibility that the transcripts containing candidates will be expressed [14].

The choices of different tissues samples are often the first consideration when study disease resistances. Liver and intestine are two major lymphoid tissues for immune response studies of several fish species. In this study, the transcriptomic profiles demonstrated that liver might be

better representative of defense responses than intestine. In contrast, intestine could contribute significantly on maintaining structure and hemostasis of intestinal barrier after infection. Firstly, when comparing with control fish, both the susceptible and resistant fish contained drastically more number of DEGs in liver than that in intestine (Table 2). Secondly, when comparing between phenotypic bulks (R vs. S), liver samples contained a much greater number of up-regulated DEGs than down-regulated ones, while intestine samples have a large fraction of down-regulated DEGs (Table 2). These resistant fish up-regulated DEGs in liver included some immune important genes, such as CC chemokine SCYA102, C-X-C motif chemokine receptor 3 (CXCR3), C-X-C motif chemokine 10 (CXCL10), toll-like receptor 13 (TLR13), TNF receptor superfamily member 1B (TNFRSF1B) and TNF alpha induced protein 8 like 2 (TNFAIP8L2). In a word, the more induced DEGs in liver than intestine, and the more up-regulated DEGs in resistant fish than in susceptible fish in liver but not intestine could be a critical indication of the phenotypic differences, thus explained why these resistant fish can survive from ESC challenge. Furthermore, gene set enrichment analysis found divergent enrichment clusters or groups between liver and intestine. Only DEGs from liver were significantly enriched into immune-related GO terms (Figure 1A). It is well recognized that immune/inflammatory responses are direct and critical processes for arousing the defense against bacteria infection. In consideration of the gene contributions to immune-related GO terms, we found majority of them were from dramatically up-regulated DEGs in resistant fish (Figure 2A). For instance, the Toll-like receptor genes (TLR7 and TLR13), CC chemokines (CXCR3, CXCL10 and SCYA102), negative regulator of reactive oxygen species (NRROS), macrophage colony-stimulating factor 1 receptor (CSF1R), TNFAIP8L2 and TNFRSF1B were

under these immune-related terms and significantly induced in resistant fish. Liver is a central immunological organ with a high exposure to circulating antigens and endotoxins from the gut microbiota, particularly enriched with innate immune cells, such as Kupffer cells, hepatic stellate cells, natural killer cells and T regulatory cells [30]. In order to maintain homeostasis, defense and tolerance mechanisms need to be activated for chronic persistence of pathogen infection, sepsis or tissue damage. In contrast, the significantly enriched functional groups from DEGs of intestine were specifically enriched on the protein degradation and biosynthesis, regulation of cell growth, transcription regulation, and wounding or wounding healing pathways (Figure 1B & 2B). Moreover, more intensive responses were observed in susceptible fish than resistant fish (Figure 2B), such as proteolysis genes coding collagenases (MMP7, MMP13 and MMP19), a disintegrin and metalloproteinase with thrombospondin motifs (ADAMTS1 and ADAMTS9), serine protease HTRA1 (HTRA1), cathepsins (CTSL and CTSD) and bone morphogenetic protein 1 (BMP1); cell growth and proliferation regulation genes coding cysteine-rich motor neuron 1 protein (CRIM1), connective tissue growth factor (CTGF), protein CYR61 (CYR61), endothelial cell-specific molecule 1 (ESM1) and insulin-like growth factor-binding protein 1 (IGFBP1); and wounding-related genes coding caveolin-1 (CAV1), hydroxycarboxylic acid receptor 1-4 (HCAR1-4), musculoskeletal, embryonic nuclear protein 1b (MUSTN1B), P2Y purinoceptor 12 (P2RY12), plasminogen (PLG) and Rho-related GTP-binding protein RhoA-C (RHOAC). Intestine serves as the first intestinal epithelium barrier caused by bacteria invasion [10]. Pathogen entry and evasion largely disrupt the structure and cellular hemostasis of intestinal tract. So the intestine would allocate more energy on the wounding or wounding healing after mechanical damage. At same time, many

damaged proteins that cannot be rescued by chaperones are needed to be degraded rapidly by proteolysis-related pathways, to avoid accumulation of cytotoxicity [50]. Therefore, general transcription and translation machineries are fairly necessary to produce new proteins for keeping intact of the intestinal barrier. Collectively, liver displays critical and essential functions in pathogen defense and tolerance mechanisms under chronic infection and sepsis. In contrast, as a primary site of infection, intestine would activate critical mechanisms in maintaining structure and hemostasis of epithelium barrier. These two tissues may collaborate closely while keep hemostasis using own specific strategies and potential unique mechanisms.

Three genomic regions were found to be associated with ESC disease in channel catfish families using the Euclidean distance equation, of which one was identified from liver SNPs and two were identified from intestine SNPs. They located in LG1, LG12 and LG26, respectively (Figure 3 & 4). These results successfully verified our previous GWAS studies and provide insights for the molecular mechanism of ESC disease resistance. The first step to process QTL signals was to identify SNPs that displayed significant differences in allele frequency using Fisher's Exact test. In this way, we can not only filter the non-relevant SNPs to the associated regions, but was able to provide initial outline for the location of candidate regions. However, since BSR-Seq works with RNA samples, the Fisher's Exact test is partially compromised by the expression levels and another source of variation into allele counts. Besides that, RNA-Seq data sets are susceptible to noise from low-coverage regions, results in false positives in estimation of potential candidates [18]. Therefore, we applied ED approach to further process the association regions. ED is a statistical matric to calculate allelic distances as the sum of differences in the allelic counts of SNPs between two different populations [34].

This method has two main advantages over traditional mapping methods, 1) the calculation metric does not require prior knowledge of parental genotypes; 2) it is linear, making it resistant to errors in allelic frequency analysis from low coverage regions [18]. Given these beneficial reasons, ED method has been well applied for identification of causal locus responsible for S-type cytoplasmic male sterility (CMS-S) in maize [35], high growth in gingko [36], seed weight in oil rape [37], black color in barley [38], multiple mutants in rice [39], and morphological phenotype in zebrafish [18]. Here, we only selected those significant SNPs for ED analysis for reducing calculation burden and redundancy instead of processing millions of SNPs. ED values of SNPs were then increased to a power 5 to enlarge the effects of candidate locus and reduce the effects low ED measurements/noise [18, 36, 39].

BSR-Seq provides a smaller candidate pool since it only reveals candidates that are expressed at the collected time points in the affected tissues [18]. Since liver and intestine play different functions and display divergent gene expression patterns following pathogen infection, it's reasonable that different QTLs should be the identified using these two tissues. After ED analyses, the QTL identified from liver were on LG1, while the other two QTLs from intestine SNPs were located on LG12 and LG26, respectively. Our recent GWAS study of ESC, resistance in channel catfish detected two significant QTL on LG1 and LG26 [6]. Remarkably, the mapped QTLs on LG1 and LG26 using BSR-Seq here were overlapped with the association regions found by GWAS (Figure 5 and 7). Such coincidence results highlighted the significance of these QTLs for ESC disease in channel catfish populations. On LG12, the QTL was not discovered by GWAS in channel catfish. However, it was identified in the fourth generation of backcross (F4 catfish) catfish populations (Figure 6), and most of the resistant alleles come

from channel catfish origin [8]. Taken together, QTLs identified by BSR-Seq successfully verified the results from GWAS studies, making BSR-Seq an efficient method for QTL mapping.

Ideally, if one genomic position is truly correlated with ESC disease resistance, it should contain allelic segregant loci and differentially expressional candidate genes. Therefore, we examined genes and their expression patterns residing within the three QTL regions. In total, eleven genes were differentially regulated between resistant fish and susceptible fish (Table 7), of which 2 genes were up-regulated and 9 genes were down-regulated. Four of these 11 genes are immune-related, two (Apo-14kDa, CCL19) were expressed higher in resistant fish than in susceptible fish, while the other two genes (Apo-Eb and NPDC1) were expressed lower in resistant fish. The gene encoding ictacalcin exhibited drastically down-regulation (40.0-fold) in resistant fish than susceptible fish. Ictacalcin was reported as a homolog of the S100 gene in channel catfish and may play function as a calcium-binding protein [40]. Here, significantly differential expression of ictacalcin in catfish liver suggests a vital role of this gene in regulating calcium homeostasis after ESC infection. The remaining differentially regulated genes are participated in various biological processes, such as ribosome biogenesis, apoptosis, cytoskeleton reorganization and protein folding (Table 7). Together, regardless of their direct or indirect involvements in disease resistance, these DEGs within QTL regions could play critical roles in maintaining hemostasis and protection against ESC disease in catfish. Further functional validation for these genes are required to better reveal their significance with ESC resistance in catfish.

A promising candidate gene should be correlated with its differential expression level and

allelic polymorphism of SNPs. With this respect, the Apo-14kDa located in the QTL of LG1 is of a great level of interest. It displayed both differentially up-regulation (Table 7) and significantly allelic segregation of SNP (Table 6) between resistant fish and susceptible fish, which strong indicate its candidacy of being involved in ESC resistance. Catfish Apo-14 kDa gene is a homologue of mammalian Apoa-II, which is one member of Apoa class [41]. In particular, Apoa and many other apolipoprotein genes were demonstrated to mediate innate immune systems. They can regulate antimicrobial activity by acting as a platform for formation of potent immunomodulatory complexes [42]. Upon infection, Apoa genes can bind to Gram-negative bacterial lipopolysaccharide (LPS) and neutralizes its toxicity through interaction between the lipid-binding domain of Apoa genes with partition of LPS [43]. As a component of the outer membrane of Gram-negative bacterial, LPS is the major mediator of Gram-negative sepsis by inducing the production of some macrophage-derived pro-inflammatory cytokines, such as tumor necrosis factor alpha (TNFa) and interleukins (IL-1a, IL-1b, and IL-6), which may lead to septic shock and, ultimately, death [44]. In mammals and teleost fish, Apoa genes were reported as potential immune modulators and antimicrobial protein, thus exhibit protective effects against sepsis [45] [46-48]. For instance, Apoa genes displayed antimicrobial activity against *Escherichia coli* in carp [46], rainbow trout [48] and striped bass [47]. It is important to note that *E. ictaluri* is also Gram-negative bacteria. It shares a number of characteristics with *E. coli,* and we already estimated that *E. ictaluri* and *E. coli* may have adopted a similar infection method [8]. Besides that, LPS is an important pathogenesis factor in *E. ictaluri,* which is useful in the design of modified vaccines against ESC [49]. Therefore, with these respects, we speculate that catfish Apo-14 kDa might have the antimicrobial activity

like Apoa-II, which could also bind to LPS of *E. ictaluri* and neutralizes its toxicity after infection, thereby inhibiting the extent of damage caused by this pathogen. Future gene function and regulation studies should be conducted to examine the antimicrobial mechanism of Apo-14 kDa and its role in innate defense against *E. ictaluri* in catfish.

In addition, a cluster of apolipoprotein genes are next to Apo-14 kDa within the QTL on LG1 and were dysregulated in liver between resistant fish and susceptible fish after infection but their differential expression was not significant (Table 8), including Apo-CI, Apo-CII, Apo-Eb, Apoa-I, and Apoa-IV. Three of these apolipoprotein genes (Apo-CI, Apo-CII and Apoa-I) were up-regulated in resistant fish that in susceptible fish while the other two genes (Apo-Eb and Apoa-IV) were down-regulated. Likewise, several apolipoprotein genes were also highlighted after ESC infection in our previous BSR-Seq study using F2 generation of backcross families, by which genes coding Apoa-IV, Apoa-II, Apo-Eb, and Apob-100 were up-regulated in resistant fish after ESC infection [11]. Such coincidence could indicate the centrality of apolipoproteins in ESC resistance/susceptibility of catfish. Many genes encoding apolipoproteins are major component of the lipid metabolism process, especially for Apoa-I, Apoa-II, Apoc-I, Apoc-II and Apo-Eb. From human and mice model, apolipoproteins have been shown to be important immune functions during infectious disease and sepsis. Except to their LPS binding and neutralization roles, apolipoprotein genes can protect against sepsis through enhancing LPS clearance, inhibit LPS-induced cytokine release and cell activation, induce an early inflammatory response during Gram-negative bacteria infection [45]. Therefore, they could play significant defense function and have great therapeutic potential for infection and sepsis.

In conclusion, in present study, we performed a state-of-art BSR-seq to study the ESC disease resistant genes in channel catfish progenies using liver and intestine tissues. Three groups were sampled as two phenotypic extremes (susceptible and resistant bulks) and one control bulk for each tissue. The methodologies of BSR-Seq allowed us to identify the DEGs among different groups, to discover significantly differentially expressed SNPs between phenotypic extremes, and finally to map QTLs underling ESC disease. Three genomic regions were found to be associated with ESC disease in channel catfish families using BSR-Seq, of which one was identified from liver SNPs and two were identified from intestine SNPs. They are located in LG1, LG12 and LG26, respectively. Together, the three identified QTLs are successfully verified our previous GWAS studies. Eleven genes were found to be differentially regulated between resistant fish and susceptible fish within the three QTL regions, suggesting their important involvements in disease resistance. In particular, Apo-14 kDa gene is most of interest because it displayed both differentially up-regulation and significantly allelic segregation of SNP between resistant fish and susceptible fish, which strong indicated that this gene could be a promising candidate for ESC resistance. All the findings in this study demonstrated that BSR-Seq is an effective and integrated solution to the disease resistance study. Further efforts still need to be conducted for elucidating gene function and involvement mechanisms for ESC resistance in catfish.

Acknowledgements

This project was supported by USDA National Institute of Food and Agriculture through a grant from Animal Disease Program (2015-67015-22975), and from Agriculture and Food

Research Initiative Animal Genomics, Genetics and Breeding Program (2017-67015-26295).

Yulin Jin is supported by a scholarship from the China Scholarship Council.

References

1. Bebak J, Wagner B: **Use of vaccination against enteric septicemia of catfish and columnaris disease by the US catfish industry**. *Journal of aquatic animal health* 2012, **24**(1):30-36.

2. Dunham RA, Elaswad A: **Catfish Biology and Farming**. *Annual review of animal biosciences* 2017(0).

3. Wolters WR, Johnson MR: **Enteric septicemia resistance in blue catfish and three channel catfish strains**. *Journal of Aquatic Animal Health* 1994, **6**(4):329-334.

4. Wolters WR, Wise DJ, Klesius PH: **Survival and antibody response of channel catfish, blue catfish, and channel catfish female × blue catfish male hybrids after exposure to *Edwardsiella ictaluri***. *Journal of Aquatic Animal Health* 1996, **8**(3):249-254.

5. Tucker CS, Hargreaves JA: **Biology and culture of channel catfish**, vol. 34: Elsevier; 2004.

6. Shi H, Zhou T, Wang X, Yang Y, Wu C, Liu S, al el: **Identification of QTL and candidate genes associated with the resistance for enteric septicemia of catfish using genome-wide association analysis in channel catfish**. *Molecular Genetics and Genomics (Accepted)* 2018.

7. Tan S, Zhou T, Wang W, Jin Y, Wang X, Geng X, Luo J, et al: **GWAS analysis using interspecific backcross progenies reveals superior blue catfish alleles responsible for strong resistance against enteric septicemia of catfish**. *Molecular Genetics and Genomics* 2018:1-14.

8. Zhou T, Liu S, Geng X, Jin Y, Jiang C, Bao L, Yao J, Zhang Y, et al: **GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance**. *Molecular Genetics and Genomics* 2017, **292**(1):231-242.

9. Peatman E, Baoprasertkul P, Terhune J, Xu P, Nandi S, Kucuktas H, Li P, Wang S, et al: **Expression analysis of the acute phase response in channel catfish (*Ictalurus punctatus*) after infection with a Gram-negative bacterium**. *Developmental & Comparative Immunology* 2007, **31**(11):1183-1196.

10. Li C, Zhang Y, Wang R, Lu J, Nandi S, Mohanty S, Terhune J, Liu Z, Peatman E: **RNA-seq analysis of mucosal immune responses reveals signatures of intestinal barrier disruption and pathogen entry following *Edwardsiella ictaluri* infection in channel catfish, *Ictalurus punctatus***. *Fish & shellfish immunology* 2012, **32**(5):816-827.

11. Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, Song L, Feng J, Liu S, Liu Z: **Bulk**

segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. *BMC genomics* 2013, **14**(1):929.

12. Michelmore RW, Paran I, Kesseli R: **Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations**. *Proceedings of the national academy of sciences* 1991, **88**(21):9828-9832.

13. Ramirez-Gonzalez RH, Segovia V, Bird N, Fenwick P, Holdgate S, Berry S, Jack P, Caccamo M, Uauy C: **RNA-Seq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat**. *Plant biotechnology journal* 2015, **13**(5):613-624.

14. Miller AC, Obholzer ND, Shah AN, Megason SG, Moens CB: **RNA-seq–based mapping and candidate identification of mutations from forward genetic screens**. *Genome research* 2013, **23**(4):679-686.

15. Liu S, Yeh C-T, Tang HM, Nettleton D, Schnable PS: **Gene mapping via bulked segregant RNA-Seq (BSR-Seq)**. *PloS one* 2012, **7**(5):e36406.

16. Trick M, Adamski NM, Mugford SG, Jiang C-C, Febrer M, Uauy C: **Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat**. *BMC Plant Biology* 2012, **12**(1):14.

17. Yates SA, Swain MT, Hegarty MJ, Chernukin I, Lowe M, Allison GG, Ruttink T, Abberton MT, Jenkins G, Skøt L: **De novo assembly of red clover transcriptome based on RNA-Seq data provides insight into drought response, gene discovery and marker identification**. *BMC genomics* 2014, **15**(1):453.

18. Hill JT, Demarest BL, Bisgrove BW, Gorsi B, Su Y-C, Yost HJ: **MMAPPR: mutation mapping analysis pipeline for pooled RNA-seq**. *Genome research* 2013, **23**(4):687-697.

19. Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y, et al: **The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts**. *Nature communications* 2016, **7**:11757.

20. Li Y, Liu S, Qin Z, Waldbieser G, Wang R, Sun L, Bao L, Danzmann RG, Dunham R, Liu Z: **Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish**. *DNA Research* 2014, **22**(1):39-52.

21. Zeng Q, Fu Q, Li Y, Waldbieser G, Bosworth B, Liu S, Yang Y, Bao L, et al: **Development of a 690 K SNP array in catfish and its application for genetic mapping and validation of the reference genome sequence**. *Scientific reports* 2017,

7:40347.

22. Martin SA, Douglas A, Houlihan DF, Secombes CJ: **Starvation alters the liver transcriptome of the innate immune response in Atlantic salmon (*Salmo salar*)**. *BMC genomics* 2010, **11**(1):418.

23. Martin SAM, Blaney SC, Houlihan D, Secombes CJ: **Transcriptome response following administration of a live bacterial vaccine in Atlantic salmon (*Salmo salar*)**. *Molecular Immunology* 2006, **43**(11):1900-1911.

24. Wynne JW, O'Sullivan MG, Cook MT, Stone G, Nowak BF, Lovell DR, Elliott NG: **Transcriptome analyses of amoebic gill disease-affected Atlantic salmon (*Salmo salar*) tissues reveal localized host gene suppression**. *Marine biotechnology* 2008, **10**(4):388-403.

25. Purcell MK, Marjara IS, Batts W, Kurath G, Hansen JD: **Transcriptome analysis of rainbow trout infected with high and low virulence strains of infectious hematopoietic necrosis virus**. *Fish & shellfish immunology* 2011, **30**(1):84-93.

26. Wu N, Song Y-L, Wang B, Zhang X-Y, Zhang X-J, Wang Y-L, Cheng Y-Y, Chen D-D, Xia X-Q, Lu Y-S: **Fish gut-liver immunity during homeostasis or inflammation revealed by integrative transcriptome and proteome studies**. *Scientific reports* 2016, **6**:36048.

27. Zhu J, Li C, Ao Q, Tan Y, Luo Y, Guo Y, Lan G, Jiang H, Gan X: **Trancriptomic profiling revealed the signatures of acute immune response in tilapia (*Oreochromis niloticus*) following *Streptococcus iniae* challenge**. *Fish & shellfish immunology* 2015, **46**(2):346-353.

28. Song X, Hu X, Sun B, Bo Y, Wu K, Xiao L, Gong C: **A transcriptome analysis focusing on inflammation-related genes of grass carp intestines following infection with *Aeromonas hydrophila***. *Scientific reports* 2017, **7**:40777.

29. Yang Y, Yu H, Li H, Wang A: **Transcriptome profiling of grass carp (*Ctenopharyngodon idellus*) infected with *Aeromonas hydrophila***. *Fish & shellfish immunology* 2016, **51**:329-336.

30. Heymann F, Tacke F: **Immunology in the liver-from homeostasis to disease**. *Nature Reviews Gastroenterology and Hepatology* 2016, **13**(2):88.

31. Invernizzi P: **Liver auto-immunology: the paradox of autoimmunity in a tolerogenic organ**. *Journal of autoimmunity* 2013, **46**:1-6.

32. Welker TL, Lim C, Yildirim-Aksoy M, Shelby R, Klesius PH: **Immune response and resistance to stress and *Edwardsiella ictaluri* challenge in channel catfish, *Ictalurus punctatus*, fed diets containing commercial whole-cell yeast or yeast**

subcomponents. *Journal of the world Aquaculture Society* 2007, **38**(1):24-35.

33. Konczal M, Koteja P, Stuglik M, Radwan J, Babik W: **Accuracy of allele frequency estimation using pooled RNA-Seq**. *Molecular ecology resources* 2014, **14**(2):381-392.

34. Schneeberger K: **Using next-generation sequencing to isolate mutant genes from forward genetic screens**. *Nature Reviews Genetics* 2014, **15**(10):662.

35. Su A, Song W, Xing J, Zhao Y, Zhang R, Li C, Duan M, Luo M, Shi Z, Zhao J: **Identification of genes potentially associated with the fertility instability of S-type cytoplasmic male sterility in maize via bulked segregant RNA-Seq**. *PloS one* 2016, **11**(9):e0163489.

36. Tang HX, Li JH, Xing SY, Du SH, Wang ZT, Liu XJ, Sun LM: **RNA-seq and bulked segregant analysis of a gene related to high growth in *Ginkgo biloba* half-siblings**. *bioRxiv* 2015:034272.

37. Geng X, Jiang C, Yang J, Wang L, Wu X, Wei W: **Rapid identification of candidate genes for seed weight using the SLAF-Seq method in *Brassica napus***. *PloS one* 2016, **11**(1):e0147580.

38. Jia Q, Wang J, Zhu J, Hua W, Shang Y, Yang J, Liang Z: **Toward identification of black lemma and pericarp gene Blp1 in barley combining bulked segregant analysis and specific-locus amplified fragment sequencing**. *Frontiers in plant science* 2017, **8**:1414.

39. Yan W, Chen Z, Lu J, Xu C, Xie G, Li Y, Deng XW, He H, Tang X: **Simultaneous identification of multiple causal mutations in rice**. *Frontiers in plant science* 2017, **7**:2055.

40. Porta AR, Bettini E, Buiakova OI, Baker H, Danho W, Margolis FL: **Molecular cloning of ictacalcin: a novel calcium-binding protein from the channel catfish, *Ictalurus punctatus***. *Molecular brain research* 1996, **41**(1-2):81-89.

41. Yang Y, Fu Q, Zhou T, Li Y, Liu S, Zeng Q, Wang X, Jin Y, et al: **Analysis of apolipoprotein genes and their involvement in disease response of channel catfish after bacterial infection**. *Developmental & Comparative Immunology* 2017, **67**:464-470.

42. Gordon SM, Hofmann S, Askew DS, Davidson WS: **High density lipoprotein: it's not just about lipid transport anymore**. *Trends in Endocrinology & Metabolism* 2011, **22**(1):9-15.

43. Henning MF, Herlax V, Bakás L: **Contribution of the C-terminal end of apolipoprotein AI to neutralization of lipopolysaccharide endotoxic effect**. *Innate immunity* 2011, **17**(3):327-337.

44. Van Oosten M, Rensen PC, Van Amersfoort ES, Van Eck M, Van Dam A-M, Brevé JJ, Vogel T, Panet A, Van Berkel TJ, Kuiper J: **Apolipoprotein E protects against bacterial lipopolysaccharide-induced lethality a new therapeutic approach to treat gram-negative sepsis**. *Journal of Biological Chemistry* 2001, **276**(12):8820-8824.

45. Pirillo A, Catapano AL, Norata GD: **HDL in infectious diseases and sepsis**. In: *High Density Lipoproteins.* Springer; 2015: 483-508.

46. Concha MI, Smith VJ, Castro K, Bastías A, Romero A, Amthauer RJ: **Apolipoproteins A-I and A-II are potentially important effectors of innate immunity in the teleost fish Cyprinus carpio**. *The FEBS Journal* 2004, **271**(14):2984-2990.

47. Johnston LD, Brown G, Gauthier D, Reece K, Kator H, Van Veld P: **Apolipoprotein AI from striped bass (*Morone saxatilis*) demonstrates antibacterial activity in vitro**. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* 2008, **151**(2):167-175.

48. Villarroel F, Bastías A, Casado A, Amthauer R, Concha MI: **Apolipoprotein AI, an antimicrobial protein in *Oncorhynchus mykiss*: evaluation of its expression in primary defence barriers and plasma levels in sick and healthy fish**. *Fish & shellfish immunology* 2007, **23**(1):197-209.

49. Vinogradov E, Nossova L, Perry MB, Kay WW: **The structure of the antigenic O-polysaccharide of the lipopolysaccharide of *Edwardsiella ictaluri* strain MT104**. *Carbohydrate research* 2005, **340**(8):1509-1513.

50. Liu S, Wang X, Sun F, Zhang J, Feng J, Liu H, Rajendran K, Sun L, et al: **RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish**. *Physiological genomics* 2013, **45**(12):462-476.

51. Andrews S: **FastQC: a quality control tool for high throughput sequence data**. 2010.

52. Kim D, Pertea G, Trapnell C, Pimentel H, Kelley R, Salzberg SL: **TopHat2: accurate alignment of transcriptomes in the presence of insertions, deletions and gene fusions**. *Genome biology* 2013, **14**(4):R36.

53. Trapnell C, Roberts A, Goff L, Pertea G, Kim D, Kelley DR, Pimentel H, Salzberg SL, Rinn JL, Pachter L: **Differential gene and transcript expression analysis of RNA-seq experiments with TopHat and Cufflinks**. *Nature protocols* 2012, **7**(3):562.

54. Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R: **The sequence alignment/map format and SAMtools**. *Bioinformatics* 2009, **25**(16):2078-2079.

55. Kofler R, Pandey RV, Schlötterer C: **PoPoolation2: identifying differentiation**

between populations using sequencing of pooled DNA samples (Pool-Seq). *Bioinformatics* 2011, **27**(24):3435-3436.

56. Turner SD: **qqman: an R package for visualizing GWAS results using QQ and manhattan plots**. *BioRxiv* 2014:005165.

57. Stanke M, Morgenstern B: **AUGUSTUS: a web server for gene prediction in eukaryotes that allows user-defined constraints**. *Nucleic acids research* 2005, **33**(suppl_2):W465-W467.

58. Pruitt KD, Tatusova T, Maglott DR: **NCBI Reference Sequence (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins**. *Nucleic acids research* 2005, **33**(suppl_1):D501-D504.

59. Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T: **Cytoscape: a software environment for integrated models of biomolecular interaction networks**. *Genome research* 2003, **13**(11):2498-2504.

60. Bindea G, Mlecnik B, Hackl H, Charoentong P, Tosolini M, Kirilovsky A, Fridman W-H, Pagès F, Trajanoski Z, Galon J: **ClueGO: a Cytoscape plug-in to decipher functionally grouped gene ontology and pathway annotation networks**. *Bioinformatics* 2009, **25**(8):1091-1093.

Table 1. Summary of RNA-Seq results of susceptible, resistant and control bulks from liver and intestine tissues after ESC infection

	Condition	Bulks	Raw reads		Clean reads		Genome mapping			Gene mapping		
			Number	Mean	Number	Mean	Number	Percentage	Mean	Number	Percentage	Mean
Liver	Susceptible	S1	46,622,922		38,771,916		31,614,956	81.20%		14,741	53.02%	
		S2	45,658,744	46,205,347	37,242,890	38,112,094	31,574,094	84.50%	83.13%	15,261	54.89%	53.98%
		S3	46,334,376		38,321,476		32,167,188	83.70%		15,053	54.14%	
	Resistant	R1	56,004,166		46,849,740		39,492,350	83.90%		14,714	52.92%	
		R2	50,042,928	48,527,543	41,191,706	40,185,901	34,857,750	84.30%	83.83%	12,748	45.85%	50.81%
		R3	39,535,536		32,516,258		27,181,312	83.30%		15,458	55.60%	
	Control	C	53,677,310		44,074,610		37,404,468	84.50%		14,175	50.98%	
Intestine	Susceptible	S1	60,015,166		48,854,364		38,187,844	77.70%		18,841	67.76%	
		S2	50,584,214	55,008,477	40,191,028	44,508,558	32,088,860	77.70%	78.53%	18,465	66.41%	67.26%
		S3	54,426,052		44,480,282		35,880,868	80.20%		18,798	67.61%	
	Resistant	R1	45,993,212		37,776,274		29,371,186	77.30%		18,318	65.88%	
		R2	37,167,544	42,962,053	30,389,018	35,367,569	23,847,940	78.00%	78.40%	17,846	64.19%	65.10%
		R3	45,725,402		37,937,416		30,768,368	80.70%		18,134	65.22%	
	Control	C	43,265,176		35,685,114		28,675,426	79.90%		18,051	64.92%	
Total			675,052,748		554,282,092		453,112,610			23,699		

Table 2. The number of DEGs from different comparisons in both liver and intestine tissues after ESC infection.

	DEGs in liver			DEGs in intestine		
Comparison*	S vs C	R vs C	R vs S	S vs C	R vs C	R vs S
Total	3092	1,652	607	2241	917	678
Up-regulated	1759	1,001	345	1123	320	277
2-5 fold	1201	459	320	897	267	252
5-10 fold	277	312	17	145	27	21
10-50 fold	245	205	8	75	26	4
>50 fold	36	25	0	6	0	0
Down-regulated	1333	651	262	1118	597	401
2-5 fold	874	375	220	976	549	346
5-10 fold	274	188	21	104	28	42
10-50 fold	168	80	20	35	20	13
>50 fold	17	8	1	3	0	0

Table 3. Information of enriched GO terms from DEGs in liver

GO terms	Count	%	P-value
GO:0055114~oxidation-reduction process	36	5.84	4.2E-05
GO:0006629~lipid metabolic process	15	2.44	8.1E-04
GO:0006954~inflammatory response	13	2.11	5.3E-04
GO:0032496~response to lipopolysaccharide	7	1.14	2.4E-03
GO:0006096~glycolytic process	6	0.97	2.5E-03
GO:0002376~immune system process	6	0.97	3.7E-03
GO:0006631~fatty acid metabolic process	6	0.97	4.6E-03
GO:0010941~regulation of cell death	3	0.65	1.1E-03

Table 4. Information of enriched GO terms from DEGs in intestine

GO terms	Count	%	P-value
GO:0055114~oxidation-reduction process	40	6.19	3.5E-06
GO:0006508~proteolysis	37	5.73	1.7E-06
GO:0000122~ transcription regulation	11	1.70	2.8E-03
GO:0001558~regulation of cell growth	9	1.39	1.3E-05
GO:0009617~response to bacterium	8	1.08	2.2E-04
GO:0006695~cholesterol biosynthetic process	3	0.62	6.8E-03
GO:0051258~protein polymerization	3	0.46	2.4E-03

Table 5. Identification of SNPs and significant SNPs from liver and intestine. The significant SNPs means the allele differences of the SNPs are significantly different between the two bulks after two-tailed Fisher's Exact test. The threshold was set as FDR p-value ≤0.05.

	Liver	Intestine
Total number of SNPs	145,377	206,328
Number of significant SNPs	10,790 (7.4%)	12,986 (6.3%)

Table 6. Information of significant SNPs with high ED[5] values.

Tissue	Linkage group	SNP location	ED[5]	Genotype	Annotation	Gene	Gene description
						Apo-14Kda	14KDa Apolipoprotein
		7,468,885	1.4	A to C	Exon_variant		
		7,491,492	2.1	G to A			
Liver	1	7,491,537	2.8	G to C	Intron_variant	Apoc-I	Apolipoprotein C-1
	12	24,564,284	1.3	T to C	Upstream_gene_variant	POLR3H	RNA polymerase III subunit H
Intestine	26	5,146,464	1.1	C to T	Intron_variant	NOXA1	NADPH oxidase activator 1

Table 7. The 11 genes with differential expression between resistant fish and susceptible fish within the three QTL regions.

Tissue	Linkage group	Gene	Gene description	Position	Fold change	Regulated (R vs S)	Biological process or function
Liver	1	Apo-14kDa	Apolipoprotein-14kda	7,467,993-7,469,576	2.5	Up	Antimicrobial activity
		Apo-Eb	Apolipoprotein Eb	7,498,727-7,501,420	2.6	Down	Immune response
		GTPBP10	GTP binding protein 10	7554768-7,563,672	3.4	Down	Ribosome biogenesis
		LOC100305037	Ictacalcin	7,863,626-7,867,236	40.0	Down	Calcium homeostasis
		EFNA1	Ephrin-A1	7,996,913-8,008,966	2.4	Down	Ephrin receptor binding
Intestine	12	XKR8	XK-related protein 8	23,221,326-23,225,051	2	Down	Apoptosis
		CYR61	Protein CYR61	24,749,123-24,759,278	2.4	Down	Extracellular matrix binding
		TNIK	Traf2 and NCK-interacting protein kinase	25,046,691-25,047,581	2.5	Down	Cytoskeleton reorganization
	26	CCL19	C-C motif chemokine 19	4,478,046-4,479,346	3.9	Up	Inflammatory response
		NPDC1	Neural proliferation differentiation and control protein 1	5,314,163-5,340,787	1.9	Down	Regulation of immune response
		DNAJB5	Dnaj homolog subfamily B member 5	6,051,041-6,060,592	2.2	Down	Protein folding

Table 8. Genes and their fold change in liver between resistant fish and susceptible fish within QTL region on LG1. The differentially expressed genes (p-value < 0.01, fold change ≥ 2) and their fold changes are in bold.

Gene	Gene description	Start	End	Fold change
LOC108265267	Uncharacterized LOC108265267	6,550,646	6,579,534	1.4
LOC108265322	Zinc finger protein Xfin-like	6,581,368	6,606,295	1.1
CMKLR1	Chemokine-like receptor 1	6,615,787	6,632,575	NA
HAS1	Hyaluronan synthase 1	6,685,627	6,689,055	-4.6
LOC108265359	Uncharacterized LOC108265359	6,700,232	6,704,553	1.3
LOC108271012	Maternal B9.15 protein-like	6,784,883	6,791,735	-2.0
KCNN4	Potassium calcium-activated channel subfamily N member 4	6,795,373	6,841,914	2.3
sraP	Serine-rich adhesin for platelets	6,847,031	6,854,331	1.4
SMG9	SMG9, nonsense mediated mrna decay factor	6,854,743	6,866,535	1.0
LOC108265441	14-3-3 protein beta/alpha-A	6,870,593	6,892,437	-1.2
PON2	Serum paraoxonase/arylesterase 2	6,904,080	6,913,305	1.2
ELMO1	Engulfment and cell motility protein 1	6,915,882	6,924,043	NA
FXYD6	FXYD domain-containing ion transport regulator 6	6,934,745	6,954,528	1.4
NFKBID	NFKB inhibitor delta	6,971,098	6,986,203	2.1
FABP7	Fatty acid-binding protein, brain	6,994,989	6,996,078	-3.4
LOC108263199	Uncharacterized LOC108263199	7,006,294	7,008,200	-1.5
LOC108265448	Uncharacterized LOC108265448	7,017,940	7,026,671	1.1
CADM4	Cell adhesion molecule 4	7,034,301	7,219,888	1.3
RPS19	Ribosomal protein S19	7,253,331	7,258,327	1.2
LOC108273183	Uncharacterized LOC108273183	7,293,420	7,295,399	-1.1
BCAM	Basal cell adhesion molecule	7,332,821	7,385,193	1.4
ZNF227	Zinc finger protein 227	7,410,177	7,414,980	1.5
LOC108259849	Uncharacterized LOC108259849	7,418,878	7,422,706	-1.6
SEC16B	Protein transport protein Sec16B	7,425,046	7,432,498	NA
APOC2	Apolipoprotein C-II	7,450,140	7,451,717	1.4
APO-14KDA	**Apolipoprotein-14kda**	**7,467,993**	**7,469,576**	**2.5**
LOC108265484	Uncharacterized LOC108265484	7,478,343	7,480,132	1.1
LOC108263427	Uncharacterized LOC108263427	7,484,762	7,486,152	-1.9
APOC1	Apolipoprotein C1	7,491,201	7,492,698	1.2
APOEB	**Apolipoprotein Eb**	**7,498,727**	**7,501,420**	**-2.6**
APOA1	Apolipoprotein A-I	7,524,057	7,529,499	1.5
APOA4	Apolipoprotein A-IV	7,545,488	7,547,051	-1.6
GTPBP10	**GTP binding protein 10**	**7,554,768**	**7,563,672**	**-3.4**
OSGIN2	Oxidative stress induced growth inhibitor family member 2	7,565,921	7,592,881	1.3

LOC108265532	Uncharacterized LOC108265532	7,594,772	7,616,151	1.7
NBN	Nibrin	7,594,772	7,616,151	1.1
FAM189B	Family with sequence similarity 189 member B	7,620,228	7,635,438	NA
RAB13	RAB13, member RAS oncogene family	7,637,617	7,649,405	-1.2
RPS27	40S ribosomal protein S27-1	7,652,722	7,660,588	-1.2
KCNN2	Small conductance calcium-activated potassium channel protein 2	7,662,799	7,762,716	-9.8
ADAR	Adenosine deaminase, RNA specific	7,783,803	7,801,533	1.6
SNX27	Sorting nexin family member 27	7,806,809	7,837,172	1.5
GTF2IRD1	General transcription factor II-I repeat domain-containing protein 2	7,844,362	7,858,984	NA
LOC100305037	**Ictacalcin**	**7,863,626**	**7,867,236**	**-40.0**
LOC108265566	Ictacalcin-like	7,872,902	7,881,164	1.1
S100B	Protein S100-B	7,898,563	7,907,387	NA
S100A1	S100-a1	7,917,738	7,919,626	-1.1
KRTCAP2	Keratinocyte associated protein 2	7,923,390	7,927,641	-1.2
TRIM46	Tripartite motif containing 46	7,929,085	7,954,395	1.4
SLC50A1	Solute carrier family 50 member 1	7,976,830	7,992,493	-1.2
EFNA1	**Ephrin-A1**	**7,996,913**	**8,008,966**	**-2.4**
EFNA3	Ephrin-A3	8,073,373	8,177,760	-2.4
LOC108265691	Uncharacterized protein LOC108265691	8,281,141	8,282,658	NA
CASPASE-1-A	Caspase-1-A	8,301,743	8,302,284	NA
LOC108265768	Uncharacterized protein LOC108265768	8,306,573	8,307,907	NA
LOC108265685	Uncharacterized protein LOC108265685	8,398,900	8,400,417	NA
LOC108265703	Uncharacterized protein LOC108265703	8,412,300	8,413,793	NA
LOC108265679	Uncharacterized protein LOC108265679	8,547,884	8,549,416	NA

Table 9. Genes and their fold change in intestine between resistant fish and susceptible fish within QTL region on LG12. The differentially expressed genes (p-value < 0.01, fold change ⩾ 2) and their fold changes are in bold.

Gene	Gene description	Start	End	Fold change
XKR8	**XK-related protein 8**	**23,221,326**	**23,225,051**	**-2.0**
KIAA0895	KIAA0895 ortholog	23,330,369	23,336,962	-1.5
CHTOP	Chromatin target of PRMT1	23,337,057	23,346,735	-1.3
THAP7	THAP domain containing 7	23,485,644	23,492,304	1.6
S100A14	S100 calcium binding protein A14	23,493,107	23,498,512	1.2
SEMA4A	Semaphorin-4A	23,550,674	23,576,115	-1.1
TAF12	Transcription initiation factor TFIID subunit 12	23,636,923	23,639,109	NA
MUC2	Mucin-2	23,642,105	23,653,897	NA
POLR3G	DNA-directed RNA polymerase III subunit RPC7	23,659,794	23,662,887	-1.6
TXNIP	Thioredoxin-interacting protein	23,691,541	23,693,721	-2.1
MTX1	Metaxin-1	23,698,899	23,711,269	-1.2
THBS3B	Thrombospondin-3b	23,720,736	23,736,381	2.4
GH1	Somatotropin precursor	23,756,162	23,758,257	NA
LRP12	Low-density lipoprotein receptor-related protein 12	23,761,204	23,769,056	-1.1
ZFPM2	Zinc finger protein ZFPM2	23,803,061	23,832,950	-2.0
CLEC4E	C-type lectin domain family 4 member E	23,836,708	23,842,254	1.0
SCN4AB	Sodium channel protein type 4 subunit alpha B	23,853,968	23,880,675	-1.1
MYL4	Myosin light chain 4	23,911,410	23,916,608	NA
MAPT	Microtubule-associated protein tau	23,933,647	23,951,005	-1.3
KAT8	KAT8 regulatory NSL complex subunit 1	23,960,369	23,995,532	NA
LOC108272857	S-antigen protein	23,976,712	23,977,143	1.4
CDC27	Cell division cycle protein 27 homolog	24,018,866	24,035,855	-1.0
SP140	Nuclear autoantigen Sp-140-like protein	24,092,314	24,147,247	1.0
AIRE	Autoimmune regulator	24,115,682	24,120,092	3.4
SP140	Nuclear body protein SP140	24,129,718	24,136,997	1.0
TRIM33	E3 ubiquitin-protein ligase TRIM33	24,171,080	24,177,600	1.3
NLRP3	NACHT, LRR and PYD domains-containing protein 3	24,211,400	24,224,975	NA

NLRP12	NACHT, LRR and PYD domains-containing protein 12	24,284,951	24,309,515	NA
POLR2A	DNA-directed RNA polymerase II subunit RPB1	24,319,018	24,320,316	NA
UBTF	Nucleolar transcription factor 1	24,481,199	24,488,878	-1.5
GUCY2C	Heat-stable enterotoxin receptor	24,490,786	24,512,010	1.2
PMM1	Phosphomannomutase 1	24,523,191	24,540,297	-1.5
CSDC2	Cold shock domain-containing protein C2	24,546,623	24,551,062	NA
POLR3H	DNA-directed RNA polymerase III subunit RPC8	24,556,976	24,561,297	-1.4
ACO2	Aconitate hydratase, mitochondrial	24,570,220	24,571,108	NA
TOB2	Protein Tob2	24,579,065	24,580,249	1.3
TEF	Thyrotroph embryonic factor	24,592,523	24,597,161	NA
ZC3H7B	Zinc finger CCCH domain-containing protein 7B	24,608,925	24,626,548	NA
RANGAP1	Ran gtpase-activating protein 1	24,629,523	24,636,705	NA
MYH9	Myosin-9	24,663,253	24,687,155	NA
TXN2	Thioredoxin, mitochondrial	24,724,942	24,729,426	1.0
CYR61	Protein CYR61	24,749,123	24,759,278	-2.4
DDX39A	**ATP-dependent RNA helicase DDX39A**	**24,766,196**	**24,774,518**	**1.0**
LRRC8D	Volume-regulated anion channel subunit LRRC8D	24,780,554	24,785,365	NA
GFI1B	Zinc finger protein Gfi-1b	24,806,010	24,809,416	-1.9
TGFBR3	Transforming growth factor beta receptor type 3	24,811,534	24,830,755	1.0
PKN2	Serine/threonine-protein kinase N2	24,833,726	24,862,232	NA
NACA	Nascent polypeptide-associated complex subunit alpha	24,881,593	24,896,312	NA
MCOLN1	Mucolipin-1	24,919,967	24,928,914	1.6
TRAPPC5	Trafficking protein particle complex subunit 5	24,930,819	24,932,917	-1.4
UBN2	Ubinuclein-2	24,935,655	24,946,849	NA
HOOK2	Protein Hook homolog 2	24,959,894	24,978,161	-1.2
TNIK	**Traf2 and NCK-interacting protein kinase**	**25,046,691**	**25,047,581**	**-2.5**
RTBDN	Retbindin	25,056,552	25,061,983	-1.2
MAST1	Microtubule-associated serine/threonine kinase 1	25,078,608	25,095,224	-1.4
ATG4D	Cysteine protease ATG4D-like	25,101,821	25,105,338	1.3
EVPL	Envoplakin	25,122,911	25,141,453	1.4

SPHK1	Sphingosine kinase 1	25,144,260	25,159,834	NA
FBF1	Fas-binding factor 1	25,165,945	25,179,710	-1.3
CYGB1	Cytoglobin-1	25,197,186	25,202,340	-1.0

Table 10. Genes and their fold change in intestine between resistant fish and susceptible fish within QTL region on LG26. The differentially expressed genes (p-value < 0.01, fold change ⩾ 2) and their fold changes are in bold.

Gene	Gene name	Start	Stop	Fold change
AIFM3	Apoptosis-inducing factor 3	4,115,486	4,133,468	-1.4
CLDN24	Putative claudin-24	4,152,825	4,153,520	-3.2
LZTR1	Leucine-zipper-like transcriptional regulator 1	4,158,743	4,172,948	1.1
POLE	DNA polymerase epsilon catalytic subunit A	4,174,058	4,200,286	NA
P2RX2	P2X purinoceptor 2	4,203,619	4,214,504	1.8
PES1	Pescadillo homolog	4,215,956	4,227,791	NA
GAL3ST1	Galactosylceramide sulfotransferase	4,239,974	4,243,552	NA
PPIL2	Peptidyl-prolyl cis-trans isomerase-like 2	4,247,907	4,284,076	NA
YPEL1	Protein yippee-like 1	4,315,528	4,333,890	NA
MAPK1	Mitogen-activated protein kinase 1	4,341,704	4,363,819	1.3
PPM1F	Protein phosphatase 1F	4,370,875	4,377,831	-1.2
TOP3B	DNA topoisomerase 3-beta-1	4,386,260	4,404,171	-1.1
SDF2L1	Stromal cell-derived factor 2-like protein 1	4,440,721	4,444,005	-1.1
YDJC	Carbohydrate deacetylase	4,448,032	4,454,042	NA
UBE2L3	Ubiquitin-conjugating enzyme E2 L3	4,458,072	4,462,728	-1.2
CCL19	**C-C motif chemokine 19**	**4,478,046**	**4,479,346**	**2.9**
TCTN2	Tectonic-2	4,482,164	4,493,847	NA
SEC14L2	SEC14-like protein 2	4,496,666	4,516,239	1.2
CNNM4	Metal transporter CNNM4	4,519,952	4,551,570	-1.2
CKAP2	Cytoskeleton-associated protein 2	4,557,509	4,570,588	-1.3
PURB	Transcriptional activator protein Pur-beta	4,596,733	4,597,650	1.2
HS3ST1	Heparan sulfate glucosamine 3-O-sulfotransferase 1	4,605,006	4,605,917	NA
BARX1	Barh-like 1 homeobox protein	4,673,845	4,676,363	NA
DDX31	Probable ATP-dependent RNA helicase DDX31	4,678,675	4,699,610	-1.2
AK8	Adenylate kinase 8	4,701,786	4,730,176	NA
TSC1	Hamartin-like	4,740,021	4,755,330	1.2

GRIN1	Glutamate receptor ionotropic, NMDA 1	4,763,636	4,795,533	-1.8
ENTPD2	Ectonucleoside triphosphate diphosphohydrolase 2	4,811,312	4,826,125	1.1
NELFB	Negative elongation factor B	4,826,922	4,835,207	1.5
LPAR6	Lysophosphatidic acid receptor 6	4,839,568	4,840,590	1.3
P2RY2	P2Y purinoceptor 2	4,867,168	4,868,187	1.5
ARRDC1	Arrestin domain-containing protein 1	4,871,238	4,882,422	-1.3
EHMT1	Histone-lysine N-methyltransferase EHMT1	4,887,407	4,903,322	1.1
CACNA1B	Voltage-dependent N-type calcium channel subunit alpha-1B	4,909,297	5,041,824	1.1
ANXA11	Annexin A11	4,931,816	4,936,620	NA
COL1A1	Collagen alpha-1(I) chain	5,046,578	5,112,401	NA
MAN1B1	Endoplasmic reticulum mannosyl-oligosaccharide 1,2-alpha-mannosidase	5,114,486	5,123,494	1.2
DPP7	Dipeptidyl peptidase 2	5,126,800	5,135,291	-1.3
NOXA1	NADPH oxidase activator 1	5,137,677	5,150,000	-2.7
WHRN	Whirlin	5,165,360	5,240,007	NA
STXBP1	Syntaxin-binding protein 1	5,245,692	5,265,159	-1.4
FAM129A	Niban	5,269,195	5,310,484	-1.4
NPDC1	**Neural proliferation differentiation and control protein 1**	**5,314,163**	**5,340,787**	**-2.2**
	Serine protease inhibitor 2.1	5,344,803	5,347,517	NA
TRAFD1	TRAF-type zinc finger domain-containing protein 1	5,360,791	5,369,832	1.0
TRAF2	TNF receptor-associated factor 2	5,372,551	5,380,844	-1.3
TNC	Tenascin	5,473,201	5,511,961	1.0
B3GALT2	Beta-1,3-galactosyltransferase 2	5,518,923	5,519,894	1.5
TESK1	Dual specificity testis-specific protein kinase 1	5,521,781	5,545,247	1.2
NSMF	NMDA receptor synaptonuclear signaling and neuronal migration factor	5,581,176	5,623,532	2.7
LTB4R	Leukotriene B4 receptor 1	5,627,521	5,628,414	1.5
AGTR1	Type-1 angiotensin II receptor A	5,658,141	5,659,031	2.0

PNPLA7	Patatin-like phospholipase domain-containing protein 7	5,682,571	5,705,946	-2.1
MRPL41	39S ribosomal protein L41, mitochondrial	5,709,599	5,710,006	1.2
DPH7	Diphthine methyltransferase	5,711,013	5,715,542	1.5
C22H9ORF142	Protein PAXX	5,716,151	5,719,146	1.1
ABCA2	ATP-binding cassette sub-family A member 2	5,721,311	5,796,805	1.3
PABPC1	Polyadenylate-binding protein 1	5,778,943	5,780,430	NA
PABPC1B	Polyadenylate-binding protein 1-B	5,802,819	5,804,330	NA
LIMK2	LIM domain kinase 2 isoform X1	5,811,420	5,845,132	1.5
PIK3IP1	Phosphoinositide-3-kinase-interacting protein 1	5,848,127	5,857,603	-1.5
THADA	Thyroid adenoma-associated protein homolog	5,861,963	5,881,909	1.3
RDH11	Retinol dehydrogenase 11	5,882,337	5,886,987	1.3
CAMK2G	Calcium/calmodulin-dependent protein kinase type II subunit gamma	5,890,331	5,958,714	1.5
SEC24C	Protein transport protein Sec24C	5,961,988	5,977,203	-1.3
TTF1	Transcription termination factor 1	5,981,375	5,986,222	-1.2
GSN	Gelsolin	5,986,550	5,991,405	-3.1
RAB14	Ras-related protein Rab-14	5,994,058	5,999,634	1.0
RIOK2	Serine/threonine-protein kinase RIO2	6,003,002	6,008,919	-1.4
EPG5	Ectopic P granules protein 5 homolog	6,011,734	6,038,278	-1.3
DNAJB5	**Dnaj homolog subfamily B member 5**	**6,051,041**	**6,060,592**	**3.6**
VCP	Transitional endoplasmic reticulum atpase	6,069,793	6,079,648	1.1
FANCG	Fanconi anemia group G protein	6,082,631	6,090,391	-1.1
EEF2	Elongation factor 2	6,093,472	6,104,189	-1.1
PIAS2	E3 SUMO-protein ligase PIAS2	6,106,803	6,114,311	1.0
ARID3A	AT-rich interactive domain-containing protein 3A	6,135,802	6,196,788	1.0

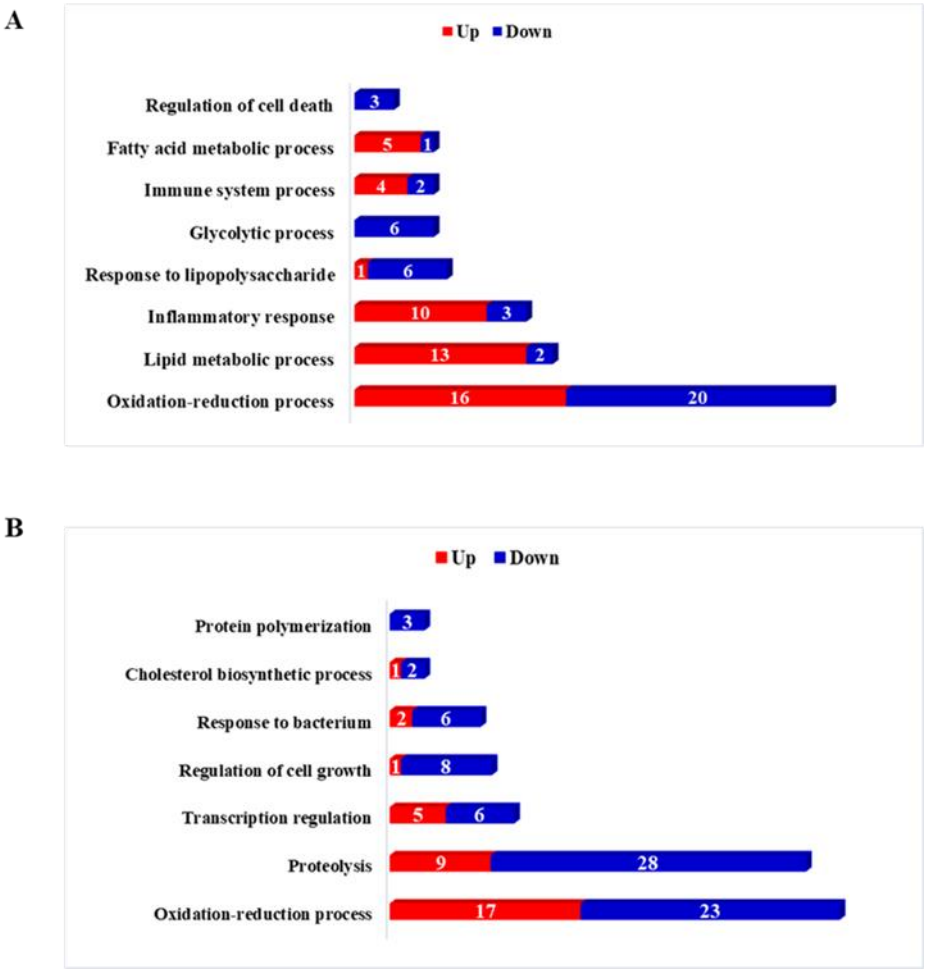

Figure 1. GO terms enrichment analysis for DEGs between resistant susceptible fish using DAVID. A. The enriched GO terms of DEGs in liver. **B.** The enriched GO terms of DEGs in intestine. Red color indicates number of up-regulated DEGs, and the blue color indicates number of down-regulated DEGs.

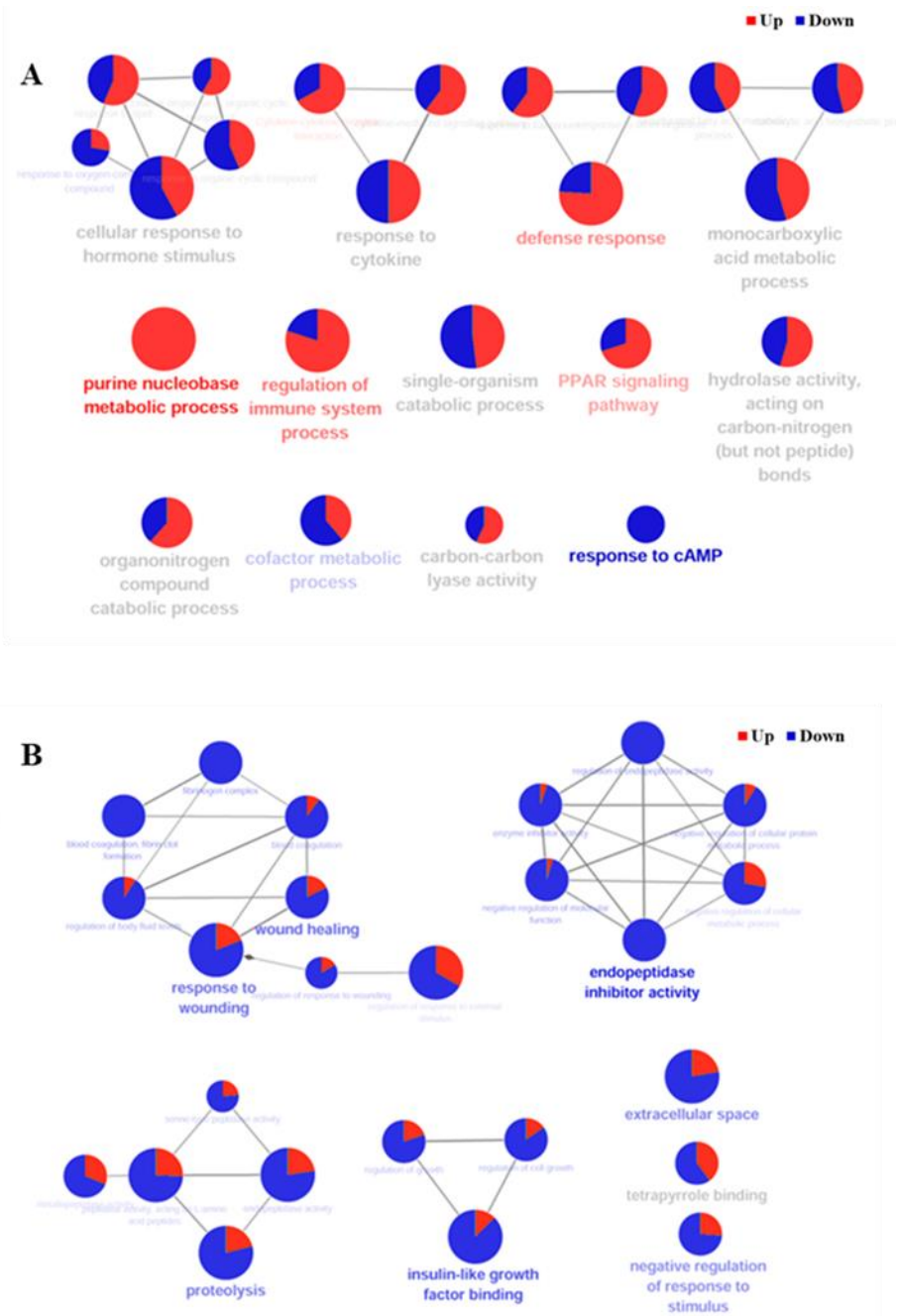

Figure 2. GO terms enrichment analysis for DEGs between resistant susceptible fish using ClueGo. A. The enriched functional groups from the DEGs in liver. **B.** The enriched functional groups from the DEGs in intestine. Red color indicates up-regulated DEGs, and the blue color indicates down-regulated DEGs. Node size represents the significance level of the enriched groups with the Bonferroni corrected p-value < 0.01.

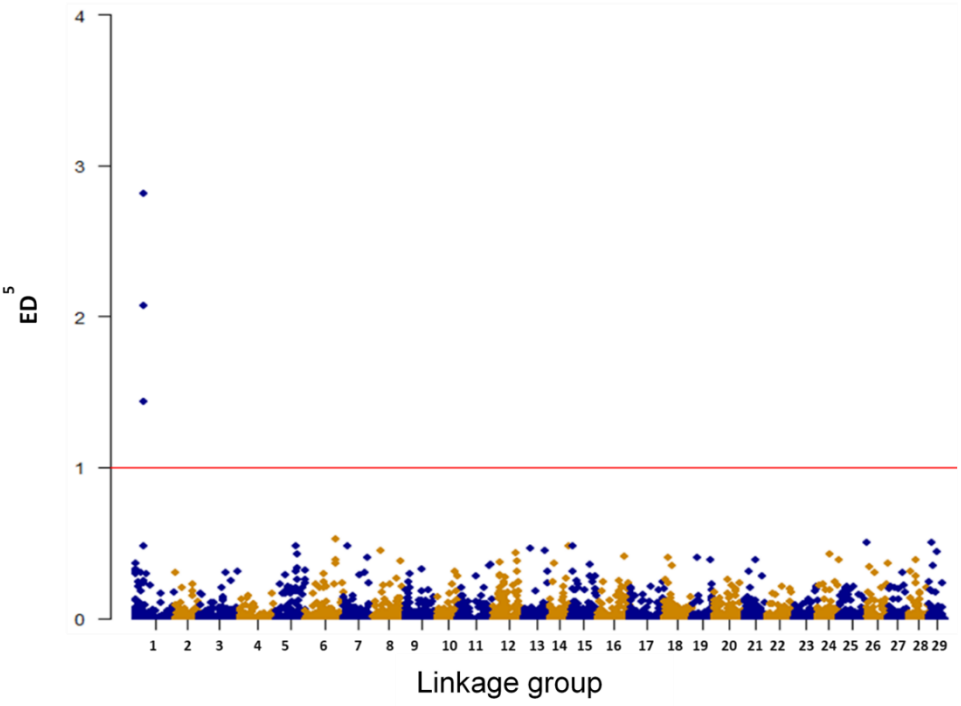

Figure 3. Manhattan plot of Euclidean distance ($ED^{5)}$ values of significant SNPs from liver for ESC disease resistance. The red solid line indicates the threshold ED^5 value for linkage region across channel catfish genome.

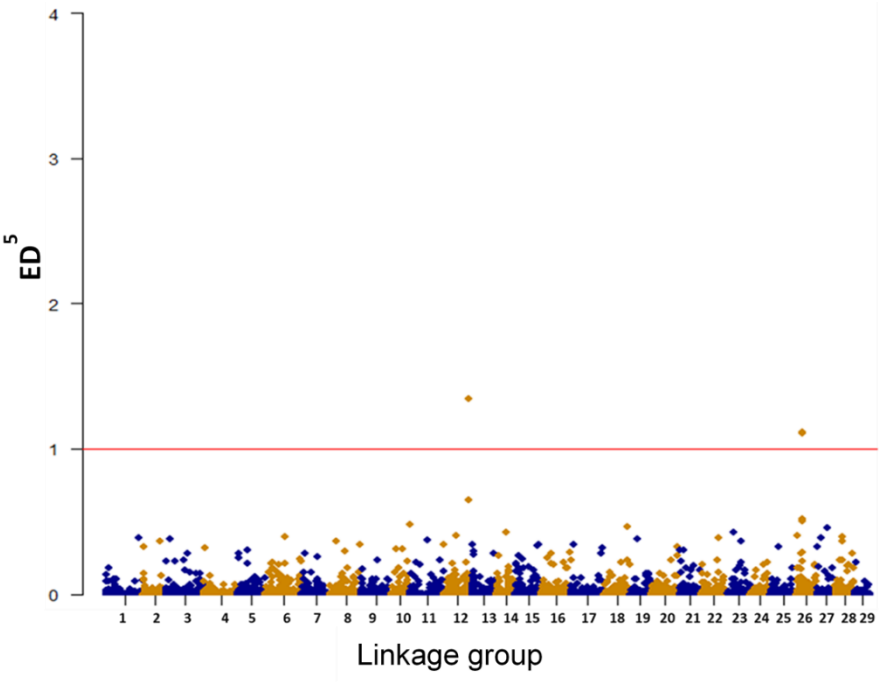

Figure 4. Manhattan plot of Euclidean distance ($ED^{5)}$ values of significant SNPs from intestine for ESC disease resistance. The red line indicates the threshold ED^5 value for linkage region across channel catfish genome.

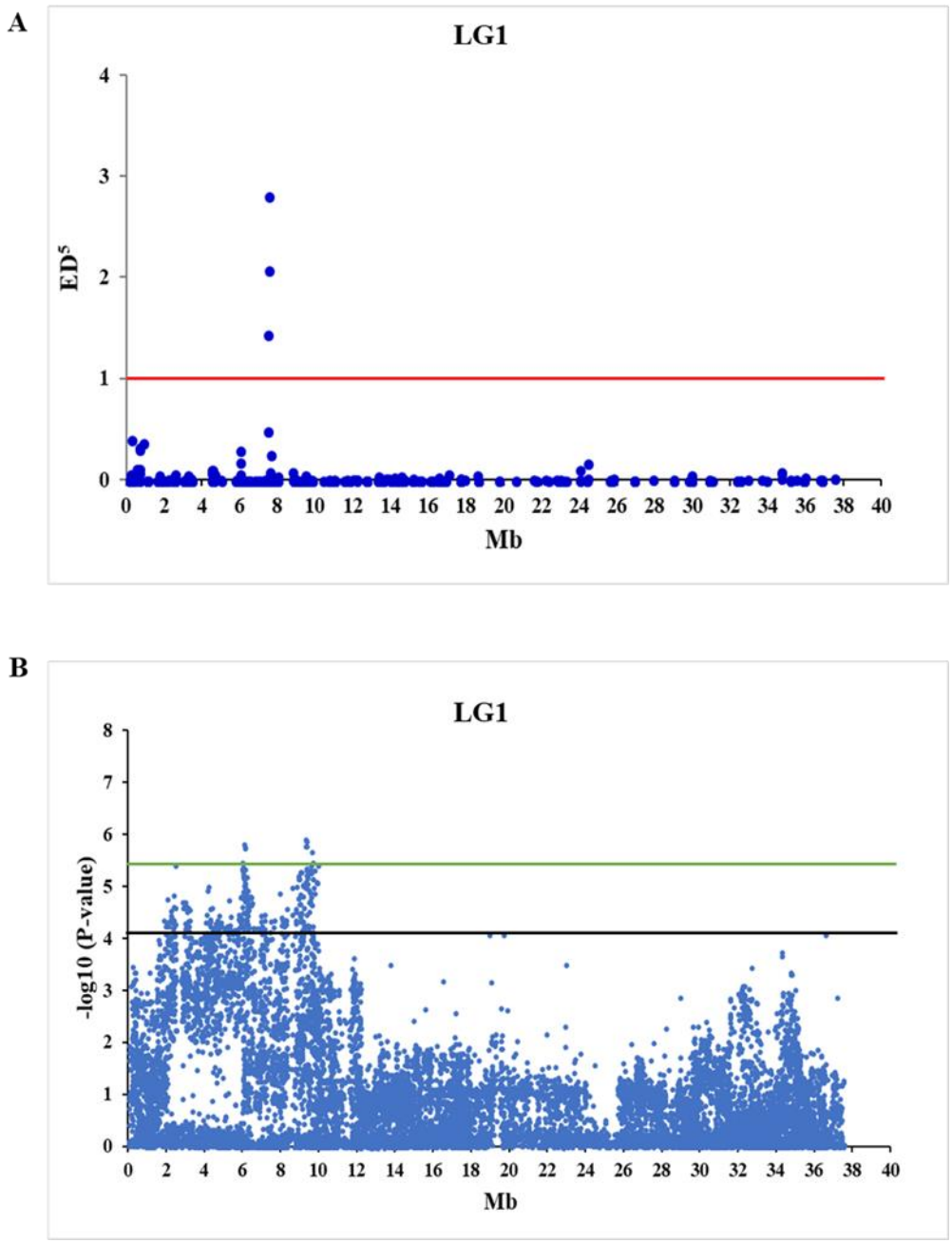

Figure 5. Comparative plots of ED^5 values and GWAS analysis for the QTL region on LG1. A. Plots of ED^5 values of significant SNPs across LG1. **B.** Plots of SNPs associated with ESC resistance on LG1 in channel catfish by GWAS [6]. Red line indicates the threshold ED^5 value for linkage region across channel catfish genome. Green solid line indicates threshold P-value for genome wide significance in GWAS. The black line indicates the threshold P value for suggestive significance in GWAS.

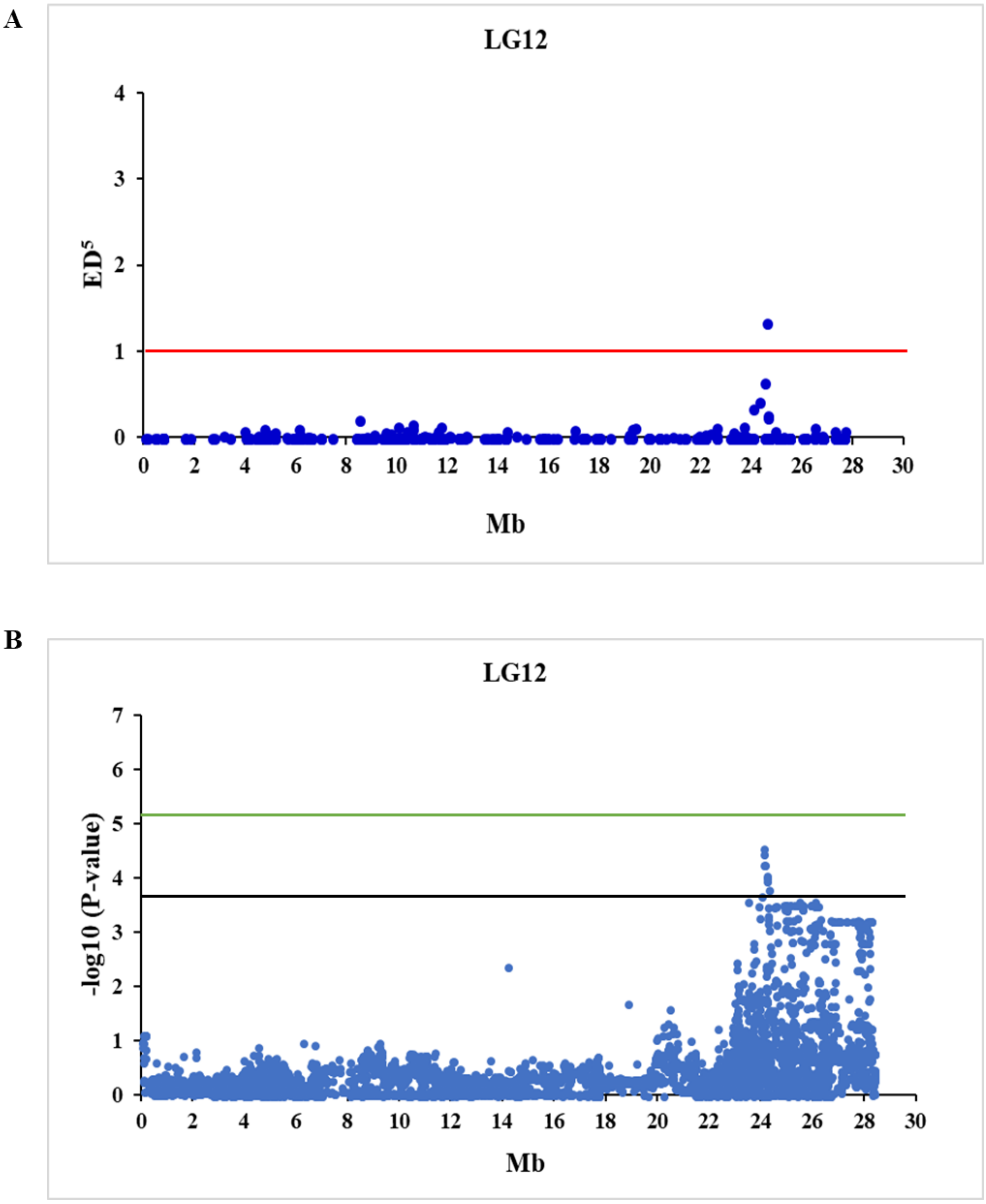

Figure 6. Comparative plots of ED^5 values and GWAS analysis for the QTL region on LG12. A. Plots of ED^5 values of significant SNPs across LG12. **B.** Plots of SNPs associated with ESC resistance on LG12 in fourth generation backcross hybrid catfish by GWAS [8]. Red line indicates the threshold ED^5 value for linkage region across channel catfish genome. Green solid line indicates threshold P-value for genome wide significance in GWAS. The black line indicates the threshold P value for suggestive significance in GWAS.

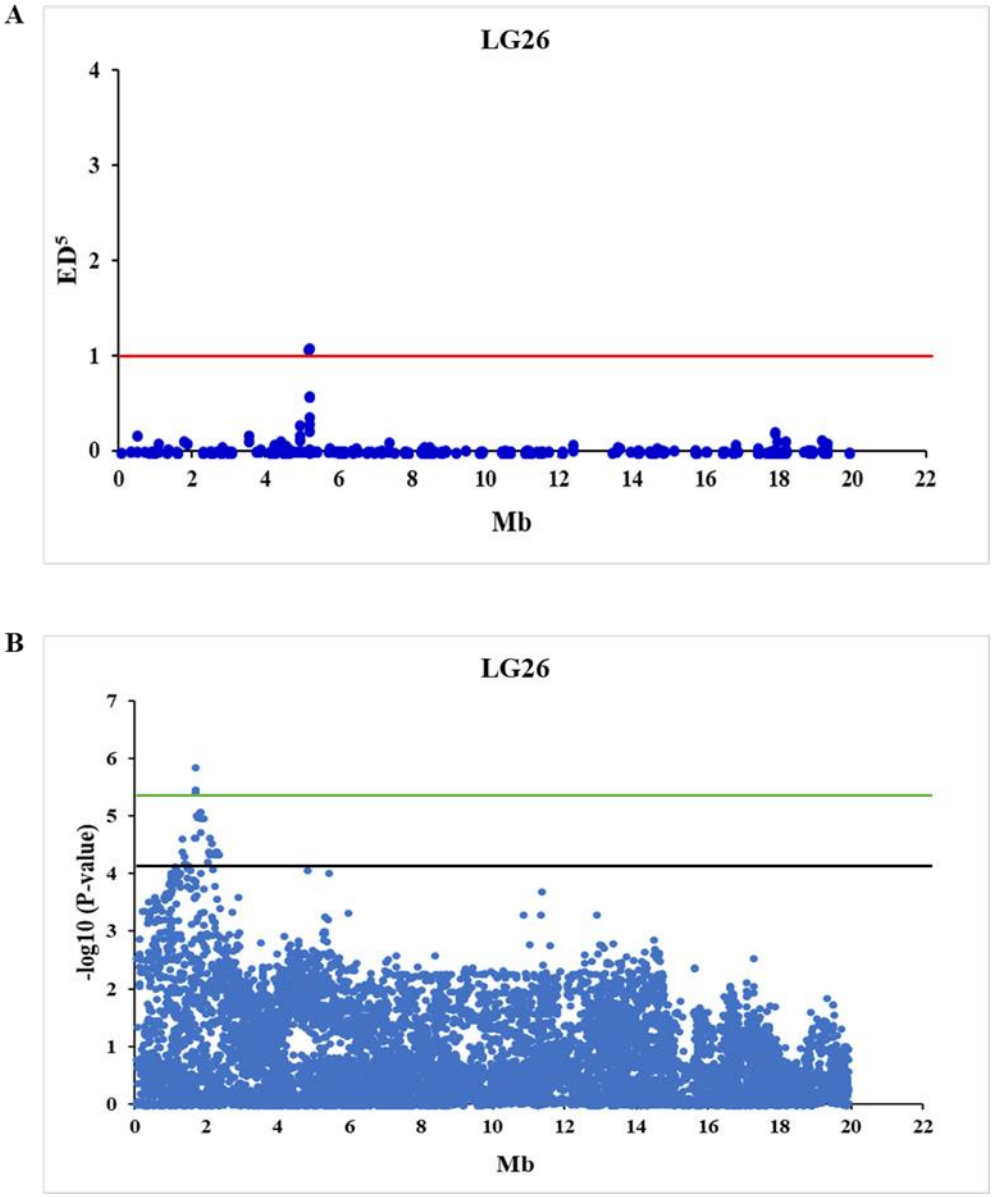

Figure 7. Comparative plots of ED^5 values and GWAS analysis for the QTL region on LG26. A. Plots of ED^5 values of significant SNPs across LG26. **B.** Plots pf SNPs associated with ESC resistance on LG26 in channel catfish by GWAS [6]. Red line indicates the threshold ED^5 value for linkage region across channel catfish genome. Green solid line indicates threshold P-value for genome wide significance in GWAS. The black line indicates the threshold P-value for suggestive significance in GWAS.