Genome-wide association studies for columnaris resistance and morphology in hybrid catfish

by

Xin Geng

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
May 8, 2016

Keywords: catfish, QTLs, GWAS, columnaris, head size

Approved by

Zhanjiang Liu, Chair, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Rex Dunham, Professor of School of Fisheries, Aquaculture, and Aquatic Sciences
Joanna Diller, Associate Professor of Department of Biological Sciences
Charles Chen, Associate Professor of Department of Crop, Soil and Environmental Sciences

Abstract

Catfish is the primary aquaculture species in the United States. Recently, the catfish industry in the USA has encountered unprecedented challenges due to increasing feed and energy costs, devastating diseases, and severe international competition. Therefore, prominent brood stocks should be developed with superior performance to profit aquaculture industry in the USA. However, little information is known about the genetic architecture controlling economically important traits, which hinders marker-assisted selection.

In this project, we studied QTLs for two economically important traits. The first trait is columnaris disease resistance. Columnaris causes severe mortalities among many different wild and cultured freshwater fish species, and it is one of the major diseases threatening catfish production. The second trait is head size (head length, head width, and head depth). Skull morphology is fundamental to evolution and biological adaptation of species to its environments. With aquaculture fish species, head size is also important for economic reasons, because it has a direct impact on fillet yield.

To identify genes associated with these economically important traits, columnaris resistance and head size, genome-wide association studies (GWAS) was performed using the catfish 250k SNP array with backcross progenies derived from crossing female channel catfish (*Ictalurus punctatus*) with male F1 hybrid catfish (female channel catfish *I. punctatus* × male blue catfish *I. furcatus*). Backcross hybrid catfish serves as a great model for the QTL analysis,

because backcross hybrids can be produced where phenotypes and genotypes are segregating, providing a useful system for QTL analysis.

The GWAS revealed four QTLs associated with columnaris resistance in catfish. Strikingly, the candidate genes may be arranged as functional hubs. The candidate genes within the associated QTLs on linkage groups 7 and 12 are not only co-localized, but also functionally related, with many of them being involved in the PI3K signal transduction pathway, suggesting its importance for columnaris resistance.

Head size QTLs were mapped in catfish to genomic regions rich in genes involved in the small GTPase pathway on nine linkage groups. Comparative analysis revealed the conserved function of small GTPase pathway in controlling skull morphometric traits in different species.

Acknowledgments

First, I would like to extend my special thanks to my advisor Dr. Zhanjiang Liu for his valuable guidance, encouragement, and support. I would also like to thank all my other committee members: Dr. Rex Dunham, Dr. Charles Chen, and Dr. Joanna Diller for their support and advice. My thanks also extend to Dr. Luxin Wang, Dr. Degui Zhi, and Dr. Peng Zeng for their support and guidance for my dissertation. I would like to thank Dr. Shikai Liu, Dr. Ruijia Wang, Dr. Chao Li, Dr. Jiaren Zhang, Dr. Luyang Sun, Dr. Yu Zhang, Dr. Yanliang Jiang, Dr. Jun Yao, Lisui Bao, Dr. Yun Li, Dr. Chen Jiang, Dr. Ailu Chen, Dr. Huseyin Kucuktas, Ludmilla Kaltenboeck, and all the other colleagues in the laboratory for their help, collaboration, and friendship. Furthermore, I would like to extend my appreciation to Dr. Tsanhai Lin, Dr. Xinyu Zhang, and the other friends in Auburn for their support in life and study. I would like to especially thank the Chinese Scholarship Council for the financial support.

Finally, I am grateful to my parents for their endless love and tremendous support.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

EMMAX	Efficient Mixed-Model Association eXpedited
GWAS	genome-wide association study
IBS	identity by state
LD	linkage disequilibrium
LG	linkage group
MAF	minimum allele frequency
PCA	principal component analysis
QC	quality control
QFAM	family-based association tests for quantitative traits
QTL	quantitative trait locus
SNP	single nucleotide polymorphism
TDT	transmission disequilibrium test

Chapter 1 Introduction

Commercial production of catfish (Ictalurus spp.) accounts for approximately 60% of US aquaculture production (www.ers.usda.gov). The catfish industry is estimated at approximately two billion dollars with value added. Catfish is one of the top agricultural commodities and is important to employment opportunities in the rural areas of the southeastern states, including Mississippi, Alabama, Louisiana, and Arkansas.

However, the catfish industry has encountered unprecedented challenges in recent years due to increasing feed and energy costs, devastating diseases, and severe international competition. For example, the cost of catfish has drastically increased recently, so profit margins have been reduced. Meanwhile, disease can cause losses of up to one third of the industry each year. Moreover, imports of catfish have risen almost 10 times in the last decade. As a consequence, the catfish industry has drastically shrunk from its peak in 2003 with 650 million pounds down to 430 million pounds in 2014 (USDA). Therefore, prominent brood stocks should be developed with superior performance.

Traditional selection breeding has been conducted for decades with aquaculture species, and major progress has been made with various traits, especially growth (Smitherman et al. 1983). However, with some traits such as disease resistance and body conformation, many genes are involved and accurate selection using traditional selection is difficult. Moreover, low heritability limits the progress of traditional selection. Whole genome marker-assisted selection

allows increased selection accuracy and efficiency, which should be conducted to develop superior brood stocks for aquaculture industry, but genetics work must be done first to dissect the genomic architecture controlling the traits of interest. More importantly, there are many important traits, and selection in one trait may adversely affect other traits, especially when the traits are closely linked. Therefore, understanding of various important performance and production traits is the prerequisite for the application of genome based selection programs.

Agricultural genetics is about the inheritance of agriculturally important traits, i.e., understanding the genetic basis of phenotypes of economic importance. The central goal of genetic stock enhancement is to discover the relationship between genetic polymorphism and the phenotypic variances observed among individuals. A phenotype of an organism is the measurement of observable characteristics or traits, while the genotype is the inherited genetic information. Qualitative traits, where the phenotypes could be assigned into different categories, are controlled by a single gene, or by a limited number of genes, such that the segregation of the traits can be followed by classical Mendelian genetics. However, quantitative traits have continuous variation, which is attributable to the combination of segregation of alleles at multiple loci controlling the trait, environment, and genotype-environment interactions. Different quantitative traits have different levels of sensitivity to genetic, sexual, and external environmental effects (Mackay 2001a). Moreover, because each causal gene may only have a small contribution to the overall heritability, identifying the genes related with quantitative traits can be difficult (Hirschhorn and Daly 2005). Most aquaculture performance and production traits of economic importance are quantitative in nature, such as grow rate, feed conversion efficiency, disease resistance for many different diseases, low oxygen tolerance, body shape, carcass and fillet yield, and behavioral traits (e.g., aggressiveness of feeding, and seinability).

Various techniques have been used to dissect the genes responsible for production and performance traits in aquaculture species. For instance, traditional QTL mapping was conducted to locate the region associated with resistance to infectious pancreatic necrosis virus in Atlantic salmon (*Salmo salar*) (Gheyas et al. 2010a; Gheyas et al. 2010b; Houston et al. 2008; Houston et al. 2012; Moen et al. 2009; Phillips et al. 2013). Utilizing RNA-seq, Li et al. (2012) characterized the role of catfish intestinal epithelial barrier following enteric septicemia of catfish (ESC, *Edwarsiella ictaluri*) challenge. Wang et al.(2013b) conducted bulked segregant analysis to study candidate gene locations and allele-specific expression associated with ESC resistance in catfish. However, genetic analysis of resistance against bacterial diseases such as columnaris has been limited.

Although many types of molecular markers can be used for marker assisted selection, SNP markers are becoming the markers of choice for two reasons. First, SNPs are abundant and widespread throughout the genomes of most species. In most aquaculture species studied to date, SNP rates are 0.5-5% among species. For instance, one SNP exists within approximately 116 bp in catfish on average (Sun et al. 2014). Such a polymorphic rate provides no limitation for a dense genome-coverage for GWAS. Although not perfectly evenly distributed, SNPs are far superior to any other types of molecular markers in these terms. Secondly, SNPs are biallelic in most cases and codominantly inherited, making them more amenable to automation with reduced complexity for genotyping and analysis.

SNPs can be readily discovered in a cost-effective fashion using the next-generation sequencing technology (Sun et al. 2014). After SNPs are defined, SNP arrays can be developed that provide high efficiency for high-throughput genotyping. Several SNP arrays have been developed for aquaculture species including catfish, carp, rainbow trout, and Atlantic salmon

(Liu et al. 2014; Palti et al. 2014; Sun et al. 2014; Xu et al. 2014). Because of these advantages, SNPs have rapidly become the marker of choice for genome-wide marker assisted selection (Morin et al. 2004).

Heterosis is an important genetic force that contributes to world food production. Interspecific hybrids are particularly effective in generating significant heterosis (Birchler et al. 2003). Hybrids of catfish have been investigated for about 50 years (Giudice 1966). Among all possible combinations, only one cross (female channel catfish X male blue catfish) exhibited better performance than its parental species (2008). The F1 hybrids made from female channel catfish X male blue catfish exhibit better performance in growth rate, feed conversion efficiency, among several other traits. Studying the mechanism of heterosis is of great economic value. Undoubtedly, hybrid catfish is the future of catfish industry, so it is economically important to identify the genetic basis underlying prominent performance of hybrid catfish.

Our long-term goal is to enhance catfish breed stocks with superior traits. In this project, we conducted GWAS for identifying QTLs controlling columnaris resistance and head size. Our specific objectives are:

1. Identifying QTLs for columnaris resistance and head size using backcross hybrid catfish with the 250K catfish SNP array to improve brood stocks by marker-assisted selection or introgression of valuable disease resistance QTLs from both channel catfish and blue catfish;
2. Identifying candidate genes, and understanding the underlying mechanisms.

Chapter 2 Introduction of GWAS

The recent breakthrough in genotyping technology allows the access to a large number of SNPs on a genomic scale. Association study, also known as linkage disequilibrium mapping, detects and locates quantitative trait loci (QTLs) based on the strength of the correlation between mapped markers and the trait in question. Although QTL mapping is well-suited for family-based samples, association studies, especially genome-wide association studies, can potentially offer higher mapping resolution using markers with higher density. Moreover, recent developments in GWAS methodologies have offered mature software packages for association analysis.

Genome-wide association study (GWAS), i.e., conducting association studies using genome-wide genotyping data, has evolved into a powerful tool for investigating the genetic architecture of important traits of human beings, crop, and animals during the last decade. For example, a genome-wide association study identifies five loci influencing facial morphology in Europeans (Liu et al. 2012). Schoenebeck et al. (2012) studied the genetic architecture of canine skull shape using GWAS, and found variation of BMP3 contributes to dog breed skull diversity. In Japanese Black cattle, a genome-wide association study identified three major QTLs for carcass weight including the PLAG1-CHCHD7 QTN for stature (Nishimura et al. 2012). 916 varieties were phenotyped under five different environments and 512 loci were identified associated with 47 agronomic traits by GWAS in *Setaria italic* (Jia et al. 2013). However, in

aquaculture, GWAS has been seldom utilized. GWAS could facilitate marker-assisted selection, and discover causative mutation. The application of GWAS in aquaculture will undoubtedly help connecting sequence diversity with phenotypic differences. In this section, a general introduction for the procedures to conduct GWAS in aquaculture is reviewed concerning the special characters of aquaculture species.

2.1 Study population

The ideal samples should be homogenous in genetic background without population stratification, highly contrasted in phenotype, and highly intercrossed to provide high mapping resolution. Population stratification is generated from the different allele frequencies among subpopulations, and it always confounds association test in practical situations. Especially when phenotypic variation exists among different subpopulations, imbalanced sampling from the subpopulations will generate false positive results. For example, if fish from strain A are more resistant to one disease than those from strain B intrinsically, it is then possible most resistant fish are sampled from strain A. As a consequence, the identified "associated" loci could be more associated with the strain difference than with disease resistance. Because the researchers are faced with various biological or economic limitations, the most appropriate samples may not be available in practical situations, which may lead to false positive results. To eliminate the effect of population stratification, a number of experimental population structures and corresponding statistical methods have been designed for GWAS. In this section, we describe a few of the most popular designs for population structure.

2.1.1 Samples from natural population

Existing samples from non-manipulated natural populations with known phenotype can be used in GWAS. Obviously, using this kind of samples is more cost- and time-effective

compared with using samples from family-based population, because the latter requires additional time to generate higher generations. While it is easy to assume natural populations are unrelated, that may or may not be true. Population stratifications could be more problematic with aquaculture species than livestocks because in many cases, a large number of individuals could be derived from a very limited number of founders, forming subpopulations in a natural population. Therefore, it must be noted that the population stratification in random natural samples could cause false positive results. Recent developments in GWAS methodologies for random natural samples have offered mature software packages for association analysis to control population stratifications, so GWAS with samples from natural population could be widely performed in aquaculture species, considering the relatively abundant natural population resources compared with livestocks and human beings.

2.1.2 Samples from family-based population

In family-based association tests, families with one or more offspring are used as the subjects rather than unrelated samples. Family-based population designs are more immune to population stratification, which cannot be efficiently addressed in natural population design. Moreover, many aquaculture species have high fecundities with thousands of progenies per spawn, saving tremendous labor for reproduction compared with livestocks. Such samples can be produced by a few parents, and the progenies with homogenous genetic background are suitable for GWAS, which cannot be realized in humans and mammals. This situation makes aquaculture species unique for GWAS using family-based samples. Even the samples consists of more than one full-sibling families in most practical experiments, the clear pedigree information of family-based population design makes correction of population stratification much easier compared with the natural population design. This and the other prominent advantages, including highly

contrasted phenotypes (sometimes realized by interspecific hybrid), allowing investigation of specific questions such as parent-of-origin effects, and power to detect rare variants, make it popular to use family-based populations for GWAS in aquaculture (Mott et al. 2000). For example, the interspecific hybrid catfish from mating female channel catfish (*Ictalurus punctatus*) with male blue catfish (*I. furcatus*) serves as a great model to detect major QTLs involved in columnaris disease resistance, because channel catfish is generally resistant to the disease while blue catfish is generally susceptible (Geng et al. 2015).

However, some disadvantages of family-based population design need to be concerned. Due to limited founders, family-based population design is not powerful to detect the causal alleles that are homozygous in the subpopulation used in the association test but heterozygous in the whole population. Moreover, compared with the natural population which has more rounds of historical recombination, the limited numbers of recombination events in family-based population design make its mapping resolution low. In addition, for aquaculture, using family-based sample requires additional breeding period. For example, the generation time of catfish is long (three years). Therefore, family based samples, especially higher generations, for association mapping are time consuming and costly. At last, the between-family stratification in both phenotype and genotype still needs to be addressed by statistical methods.

Mott et al. (2000) proposed that higher generations of intercross hybrids can be produced by intermating F2 individuals for several generations, and such higher generations of hybrids can provide a higher resolution for association mapping. They applied multi-parent advanced generation intercross (MAGIC), and the MAGIC approach provides ideal samples with high diverse and no population stratification structure, which is suitable for fine mapping. The idea of higher generations of intercross hybrids is very simple: basically, the haplotype blocks become

shorter and shorter surrounding the gene of interest to allow the identification of the candidate genes within a small chromosome region. However, in spite of the advantages and theoretical attractiveness, this approach has limited application potential for many important aquaculture species, simply because of the long generation time of aquaculture species. It takes too long to produce high enough generations of progenies to effectively reduce the linkage disequilibrium (LD) block sizes.

2.2 Phenotype design

A good understanding of the observational data and a correct adjustment for the phenotype are key prerequisite steps for further analysis. Based on the phenotypes, i.e., qualitative and quantitative traits, two types of study design can be made: qualitative trait design (case-control design) and quantitative trait design. In some cases, if the trait does not have well-established quantitative measures, the samples can be classified as categorical variable. For instance, in the case of disease resistance, the quantitative measurements may not be available for many species. Then the disease resistance trait can be classified as "resistant" versus "susceptible" as a binary variable (Geng et al. 2015). If the traits are binary, the data could be analyzed with logistic regression models. Although some methods for association tests were developed with quantitative traits, they can also be used to analyze case-control datasets by using dummy variables (i.e., coding case phenotypes as 1 and control phenotypes as 0) (Kang et al. 2010). From the statistical perspective, genetic effect size (the proportion of phenotypic variance explained by two alleles at a locus) can be easily calculated with quantitative traits, since the quantitative traits are measured by continuous numbers.

Factors that influence the trait should be adjusted in or before the association tests to exclude spurious associations caused by confounding factors. These factors, which describe the

circumstances under which the data were collected and the characteristics of the samples, may include gender, age, experimental batch, body weight, known family structure, etc. The adjustment procedure could be conducted by linear models with the factors as the explanatory variables and the observational data of interest as the response variables, and the residuals could be used as the phenotype for further analysis, which just consist of the genetic component, other unaccounted effects, and random effects (Dominik 2013). In addition, outliers should be removed because they will affect the fitting of the model.

2.3 Power of association test and sample size

Power of a study is the probability that a true association between a marker and the trait of interest is found significant by the designed study. Calculating power before conducting experiment is a central element in a study design. Power depends on the significance level α set by the experimenter, design of experiment, statistical test, effect size of QTL, the allele frequency of the causal allele, the LD between the causal allele and the genotyped markers on the array, and sample size (Hayes 2013). Increasing the sample size is an obvious method to improve the power to detect associations.

For the quantitative trait designs, selective genotyping is often utilized as a cost-effective strategy, which just genotypes individuals from the extremes of the phenotypic distribution (Lynch and Walsh 1998). It requires less sample size but keep the high power to detect QTLs (Van Gestel et al. 2000). However, the tradeoff is that it may cause the potential overestimation of effect size.

Some software are available to calculate the sample size for unrelated individuals to ensure sufficient power (Gauderman and Morrison 2006). However, including a statistician during planning phase is often recommended to ensure a solid and powerful design.

2.4 Quality control procedures

After genotype calling procedure based on signal intensities generated by the SNP assay for the alleles (Ziegler et al. 2008), quality control (QC) should be performed for genotypes to avoid false results. Quality control for GWAS data includes sample-level QC and SNP-level QC.

Samples with low genotyping quality or a low call rate should be excluded from analysis. The "outliers" with different ancestry may cause false positive loci. For example, in the study conducted by Gudbjartsson (Gudbjartsson et al.), the individuals have large deviations in terms of genetic background were removed to keep the samples homogenous. Principal component analysis or cluster analysis based on IBS (identity by state) kinship matrix with the genotypes of all samples can be used to detect outliers. After visualizing these structures, the outliers can be identified and removed (Geng et al. 2015).

For markers, SNPs with low genotyping quality should also be excluded if they have any Mendelian inheritance errors or low calling rate. The rare SNPs, possibly generated by genotyping errors or population stratification, may lead to spurious results. Therefore, SNPs with low minor allele frequencies (MAF) should always be discarded. Hardy–Weinberg equilibrium test compares the observed proportion of the marker versus the expected proportion. If unrelated samples are used, the SNPs severely out of Hardy–Weinberg equilibrium should be flagged before further analysis, because disequilibrium can result from a true association, a potential genotyping error, or population stratification (Turner et al. 2011).

2.5 Linkage disequilibrium analysis

In population genetics, linkage disequilibrium (LD) describes the correlations of alleles at two or more neighboring loci (Reich et al. 2001). If one locus has alleles A and a with frequencies p_A and p_a, and a second has alleles B and b with frequencies p_B and p_b, then the

expected haplotype frequencies at equilibrium are the product of the two component allele frequencies. For example, $p_{AB}=p_A \times p_B$, where p_{AB} is the frequency of $_{AB}$ haplotype. The deviation of the observed frequency of a haplotype from the expected under equilibrium is the linkage disequilibrium which is denoted by D: $D=p_{AB}-p_A \times p_B$ (Table 1). The D statistic is dependent on the frequencies of the individual alleles (p_A and p_B), so D is not useful in describing the LD on different pairs of loci. Two alternative methods D' and r^2 are used to normalize D (Lewontin 1964; Pritchard and Przeworski 2001):

1) $D' = \frac{D}{D_{max}}$, where $D_{max} = \begin{cases} \min(p_A p_B, p_a p_b) \; when \; D < 0 \\ \min(p_A p_b, p_a p_B) \; when \; D > 0 \end{cases}$

2) $r^2 = \frac{D^2}{p_A p_B p_a p_b}$

Table 1 Relationship between the haplotype frequencies, allele frequencies, and D.

	A	**a**	**Total**
B	$p_{AB}=p_A p_B+D$	$p_{aB}=p_a p_B-D$	p_B
b	$p_{Ab}=p_A p_b-D$	$p_{ab}=p_a p_b+D$	p_b
Total	p_A	p_a	1

LD is caused by the lack of recombinations breaking the linkage of nearby loci. Therefore LD decays with increasing distance between loci. The LD decay is also influenced by several other factors, such as population size, the number of founders in the population, and the number of generations of the populations (Bush and Moore 2012). A significantly associated

SNP detected from association mapping could be a causal variant, but in most cases, the identified SNPs are in high linkage disequilibrium with the causal variants.

There are several reasons why LD is interesting. Firstly, the resolution of the association mapping depends on the decaying extent of LD. Secondly, we could generate independent SNPs which are not correlated with the surrounding SNPs (LD pruning). Using the number of independent SNPs, we can conduct the Bonferroni correction (Geng et al. 2015).

2.6 Association test

There are different kinds of association test models. The proper statistical test method should be chosen carefully according to specific situations, for example, quantitative trait studies versus qualitative trait studies, samples from family-based population versus samples from natural population, and different genetic effects including dominant, additive and recessive. If no population stratification exists in the samples, it is simple to evaluate the association between markers and trait by common methods, including linear model, Cochran–Armitage trend test, etc. However, population stratification almost always exists within the sample population and, therefore, correction of population stratification is the key issue in association test.

Various software packages have been developed for statistical analysis of GWAS in different situations. Most are free that can be downloaded from the Internet. For example, Purcell et al. (2007) developed PLINK to conduct association test with free access, and it has been widely used in GWAS. Some commercial software packages assemble popular methods into an easy to use toolset with user-friendly interface.

In the following, the strategies will be elucidated to detect population stratification and infer genetic ancestry. Two main approaches, mixed linear model and transmission disequilibrium test, are introduced.

2.6.1 Principal component analysis

The principal components analysis (PCA), a dimensionality reduction method, is a powerful way of representing the genetic relationship. PCA summarizes the variation among different samples across all independent markers into a smaller number of principal components, which indicates the relationship of the individuals. When using PCA to correct the population stratification, the regression of genotype at a candidate SNP to the phenotype is adjusted by including the loadings of top principal components to remove all correlations to ancestry (Price et al. 2006). This method assumes a small number of ancestral populations and simple admixture, and it cannot correct stratification due to complex relationship (Yu et al. 2005). Among PCA-based software packages that have been proposed, EIGENSTRAT is the most widely used (Price et al. 2006).

2.6.2 Linear mixed models

Linear mixed model can be used to model population structure, family structure and cryptic relatedness (Yu et al. 2005). It has the ability to capture multiple levels of population structure of the samples, even from several families or inbred lines (Kang et al. 2008).

The model is listed as follows:

$$\mathbf{Y} = \mathbf{Xb} + \mathbf{a} + \mathbf{e}$$

where **Y** is the vector of phenotype; **X** is data matrix of fixed effects ; **b** is the coefficient of fixed effects; **a** is the vector of random effects with covariance structure based on kinship matrix **G** **(**Var (**a**) = $\sigma_g^2\mathbf{G}$, σ_g^2 represent the parameter for additive genetic variance); **e** is the vector of random residuals. This method models phenotypes using a mixture of fixed and random effects. Fixed effects (**Xb**) include the SNPs and optional covariates, and random effects include heritable (**a**) and non-heritable random variation (**e**) (Price et al. 2010). An efficient mixed-model

association method, EMMAX, was developed that markedly reduced the computational cost and have been widely used in GWAS (Kang et al. 2010). First, EMMAX computes a genetic relatedness matrix representing the sample structure, whose entries are genetic relationship between every pair of individuals. Second, using a variance component model, the contribution of the sample structure to the covariance of phenotype is estimated, generating an estimated covariance matrix of phenotypes that models the effect of genetic relatedness on the phenotypes. Third, a generalized least square F-test or a score test is applied at each marker to detect associations accounting for the sample structure using the covariance matrix (Kang et al. 2010). Although EMMAX was designed preferably for quantitative traits that follow a normal distribution, the association test for qualitative traits can be approximately conducted using 0-1 quantitative response variable to represent the case-control status (Kang et al. 2010). There are some similar methods, such as GCTA, TASSEL and GEMMA, which are proven to be effective in correcting complex structure stratification (Yang et al. 2011; Zhang et al. 2010; Zhou and Stephens 2012). Despite the advantage that linear mixed model could help eliminate false positives caused by complex population stratification, it is not guaranteed to adjust for all possible confounding population structures. A recommended practice is both using principal components as fixed effects and using estimated kinship matrix as variance-covariance matrix in the random effects (Price et al. 2010). Considering imperfect adjustments, the samples with less population stratification are still preferred to avoid spurious results.

2.6.3 Transmission disequilibrium test and derivatives

The transmission disequilibrium test (TDT), in which family pedigrees of samples are ascertained, is robust to the effects of population stratification (Laird and Lange 2006). The TDT was proposed by Spielman et al. (1993) with family-based populations for the association test

between a genetic marker and a trait. When conducting TDT, the progenies in each family with a certain extreme phenotype of interest are selected, for example, the fish with albino or with an exceptionally high growth rate (Lange et al. 2002). Parents and progenies are genotyped and the loci where parents are heterozygous will contribute to the analysis. From each parent, one allele must be transmitted to the progeny and the other one not. Over all families, the ratio of transmission to non-transmission will be compared with the expected value of 1:1 (Mackay and Powell, 2007).

Various extensions of the TDT have been developed, of which the family-based association test for quantitative traits is widely used (Abecasis et al. 2000). It can accommodate nuclear families of any size, with or without parental information. It breaks down the genotypes into between-family and within-family components, and the latter is free of population structure.

The major drawback of TDT is its extreme susceptibility to genotype errors of parents. TDT, which needs additional genotype information of parents, is of lower power as it only uses the allele transmission information within pedigrees. Moreover, the family based studies still need to incorporate between-family information, which may be confounded from stratification (Lasky-Su et al. 2010; Won et al. 2009).

However, with a large number of progenies, aquaculture species may be ideally suited for the TDT design. Compared with the trio-design in humans, larger family design is possible with most aquaculture species, and this advantage greatly reduces the efficiency penalty for genotyping parents. Moreover, the parental genotypes could be validated to correct the genotype errors by the genotypes of numerous offspring based on the Mendelian laws of inheritance. Considering the immunity to population stratification, family-based design will perform efficiently for most aquaculture species.

2.7 Significance level for multiple testing

Using a strict significance level for GWAS is important, because GWAS typically tests a very large number of hypotheses and spurious false positive results may arise by chance. In GWAS, the null hypothesis refers to the statement that no association exists between the markers and the trait. Thus rejecting the null hypothesis means an association. Under the null hypothesis, low P-value indicates that the chance for obtaining the observed sample results is small. When P-value falls below a predetermined alpha value (significance level), which is usually 0.05 for single marker testing, the null hypothesis will be disproved. This also means that the null hypothesis will be disproved with a probability of 5% when it is true in fact (type 1 error), so the probability for a false positive in one single test will be 5%. However, when we conduct a multiple test in GWAS, hundreds of thousand SNPs are tested simultaneously. Therefore, the cumulative likelihood of false positive results will increase. To control the false positive results, Bonferroni correction converts $\alpha=0.05$ to $\alpha=0.05/n$, where n equals the number of independent tests. Because of linkage disequilibrium among GWAS markers, each association test of all the markers is not independent. Duggal et al. (2008) proposed that the threshold P-value (α value) for genome-wide significance could be calculated based on Bonferroni correction with the estimated number of independent markers and LD blocks. For instance, if the probability of one type 1 error should be controlled at 0.05 with a total of 15,000 haplotype blocks, the genome-wide significance level now is at 0.05/15,000 = 0.0000033. Apart from Bonferroni correction, an alternative method to adjust α value is using false discovery rate (FDR), which is widely used in multiple hypothesis testing but less common in the GWAS context (Hochberg and Benjamini 1990). At last, using the $-\log_{10}$(P-value) and the positions of SNPs, Manhattan plots could be generated to show the locations of associated SNPs.

2.8 Comparison of GWAS with alternative designs

Apart from GWAS, alternative statistical methods are available to investigate the genetic basis of variation causing different phenotypes. Here, the advantages and disadvantages of these methods compared with GWAS are described.

Similar to GWAS, quantitative trait locus (QTL) mapping is also a statistical method that links phenotypic data and genotypic data to explain the genetic basis that causes phenotypic variations. QTL mapping can be regarded as a special case of GWAS where LD is derived from a small number of founders that established the population in the recent past (Mackay and Powell 2007). Different from GWAS, QTL analysis requires two or more strains of organisms as the parental population that differ genetically with regard to the trait of interest. Moreover, genetic markers which are different in the parental lines should segregate with the contrasted phenotype.

QTL mapping has been a powerful traditional method used to identify loci co-segregating with a given trait. Without significant investment in the development of large genotyping platforms such as SNP arrays, QTL mapping is still widely used. However, QTL mapping suffers from some fundamental limitations. First, the mapping resolution of QTL is limited by the amount of recombination events within the pedigrees, although it can be improved by several generations of intercrossing (Darvasi and Soller 1995). Linkage analyses thus have a lower level of resolution than association studies, which also leverages all historic recombination events among founders (Mackay 2001a). In the natural populations that are utilized by GWAS, LD often decays more rapidly with increasing physical distances than in controlled crosses (Mackay and Powell 2007). Secondly, some loci will remain undetected if the analyzed families contain no segregating alleles at the loci. Thirdly, linkage analysis have less power to identify

common genetic variants with modest effects (Risch and Merikangas 1996). Fourthly, important quantitative traits usually have complex genetic architectures, such that the phenotype is determined by multiple factors (Wang et al. 2005), such as genotype-by-sex, genotype-by-environment, and epistatic interactions between QTLs. However not all QTL studies were designed to detect such interactions (Mackay 2001b). Moreover, the allele frequencies and combinations present in the sampled families may differ from those in the other populations (Korte and Farlow 2013). Because of these reasons, the number of times that individual genes have been identified utilizing a QTL mapping remains very small.

The basic idea of bulk segregant analysis (BSA) is that phenotypic extremes should have drastic differences in the loci associated with the phenotype when samples are selected from phenotypic extremes and their genotypes are analyzed in bulk. Although the associated loci may be difficult to be detected comparing individuals with different performance in phenotype, the pooled samples (bulk) with the phenotypic extremes should reveal the contrast in the genotype (Michelmore et al. 1991; Wang et al. 2013b). In other word, if samples are grouped according to the contrasted trait, the frequency of the two marker alleles present within each of the two bulks should deviate significantly from the expected ratio in their specific population (Quarrie et al. 1999). Thus, the correlation between genotype and phenotype can be identified. The major drawback of BSA is the imprecision caused by the genotype generated from the pooled sample. Moreover, the family stratification, if existing, is impossible to be eliminated due to the bulk analysis. Further, BSA will only be able to detect the genetic effects of single locus, precluding any analysis of haplotype or gene-by-gene interaction effects. However, because of high efficiency, low cost, and analytical simplicity, it is still broadly used, especially with plant

species. The high fecundities of aquaculture species make BSA potentially a useful tool to provide preliminary result for aquaculture species (Wang et al. 2013b).

2.9 Conclusions

Undoubtedly, GWAS has accelerated the field of human, plant and livestock genetics. Using GWAS, numerous genetic risk factors for many common human diseases have been identified, and many genetic regions controlling important economical traits have been located in plants and livestocks. Genome-wide association studies could open new frontiers in our understanding of the relation of traits and the underlying genetic architecture in aquaculture. With the development of genotyping technologies, especially high-density SNP arrays, GWAS could be widely used for the analysis of aquaculture traits to improve the brood stocks of aquaculture species, with lower costs in the long term.

Chapter 3 A Genome-wide association study in hybrid catfish for QTLs controlling columnaris disease resistance

3.1 Abstract

Columnaris causes severe mortalities among many different wild and cultured freshwater fish species, but understanding of host resistance is lacking. Catfish, the primary aquaculture species in the United States, serves as a great model for the analysis of host resistance against columnaris disease. Channel catfish in general is highly resistant to the disease while blue catfish is highly susceptible. Backcross hybrids can be produced where phenotypes and genotypes are segregating, providing a useful system for QTL analysis. To identify genes associated with columnaris resistance, we performed a genome-wide association study (GWAS) using the catfish 250k SNP array with 340 backcross progenies derived from crossing female channel catfish (*Ictalurus punctatus*) with male F1 hybrid catfish (female channel catfish *I. punctatus* × male blue catfish *I. furcatus*).

A genomic region on linkage group 7 was found to be significantly associated with columnaris resistance. Within this region, five have known functions in immunity, including *pik3r3b, cyld-like, adcyap1r1, adcyap1r1-like*, and *mast2*. In addition, 3 additional suggestively associated QTL regions were identified on linkage groups 7, 12, and 14. The resistant genotypes on the QTLs of linkage groups 7 and 12 were found to be homozygous with both alleles being derived from channel catfish. The paralogs of the candidate genes in the suggestively associated

QTL of linkage group 12 were found on the QTLs of linkage group 7. 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. Strikingly, the candidate genes may be arranged as functional hubs; the candidate genes within the associated QTLs on linkage groups 7 and 12 are not only co-localized, but also functionally related, with many of them being involved in the PI3K signal transduction pathway, suggesting its importance for columnaris resistance.

3.2 Background

Flavobacterium columnare, a Gram-negative bacterium, is the causative agent of columnaris disease, which is very common in wild and cultured freshwater fish worldwide (Plumb et al. 2011). This pathogen can infect a variety of fish species through mucosal attachment points on the gill and skin, causing external erosion and necrosis (Declercq et al. 2013). The bacterium can also enter the blood stream and invade the internal organs (Hawke and Thune 1992). The economically important foodfish channel catfish (*Ictalurus punctatus*) and the other members of the family Ictaluridae are extremely susceptible to columnaris disease. Columnaris disease is considered as one of the most important diseases in the catfish industry, causing tens of millions of dollars in losses every year (Declercq et al. 2013).

There is no efficient method currently to control the disease problems in catfish other than genetic stock enhancement. Vaccines for columnaris are available now, but they are not cost efficient, because fish has low individual value compared to livestock. Additionally, the large numbers of fish make it difficult to apply vaccine to each fish. Antibiotics can be useful to control bacterial disease. However, antibiotics could have drastically adverse environmental impact and serious human health risks. Therefore, considering the moderate heredity of

columnaris disease resistance (about 0.2), improving disease resistance by genetic stock enhancement using various approaches including strain selection, crossbreeding, hybridization, and transgenics is required (Arias et al. 2012).

In the last decade, various studies were conducted aiming at elucidating the mechanisms of columnaris entry, immune evasion, and the host response to the disease. Some studies have been conducted on the modes and dynamics of columnaris adhesion (Decostere et al. 1999; Olivares-Fuster et al. 2011). In recent years, the application of RNA-seq has allowed a significantly greater level of understanding of the complexities of columnaris induced gene expression (Sun et al. 2012). Several central signatures following infection were revealed by gene expression enrichment analysis and gene pathway analysis of differentially expressed genes (Sun et al. 2012). For instance, Beck (2012) revealed that rhamnose-binding lectin was induced dramatically after infection, which was correlated with columnaris susceptibility. Peatman (2013) carried out RNA-seq analysis to compare basal and post-challenge differences in expression between susceptible and resistant channel catfish lines. Some genes involved in critical innate immunity, such as iNOS2b, lysozyme C, IL-8, and TNF-alpha were constitutively expressed higher in resistant than in susceptible catfish gill tissues. In contrast, secreted mucin forms, rhamnose-binding lectin, and some mucosal immune factors were found to be expressed at higher levels in the susceptible catfish line than in the resistant line. Despite these efforts, the knowledge of molecular mechanism of columnaris resistance is still limited.

In spite of very rapid progress made in genetic enhancement of aquaculture species with growth related traits, selection with disease resistance traits has lagged behind. To identify the QTLs in the hybrid catfish related to columnaris resistance, a genome-wide association study

using backcross hybrid catfish was conducted, and here we report the identified QTLs and their associated genes within the highly associated genomic regions.

3.3 Methods

3.3.1 Ethics statement

All experiments involving the handling and treatment of fish were approved by the Institutional Animal Care and Use Committee (IACUC) at Auburn University. Tissue samples were collected after euthanasia. All animal procedures were carried out according to the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States.

3.3.2 Experimental fish, bacteria challenge and sample collection

The study population was the Auburn University one year old catfish generated from back cross of male F1 hybrid catfish (female channel catfish X male blue catfish) with female channel catfish. The backcross progenies were produced by using the F1 as the male parent to avoid possible maternal effects in the early growth stage. The population consisted of six families. Since the offspring were mixed for culture, the genotypes of the samples could be used to assign the offspring into their families (see Table 2). Three reasons make us to choose the backcross family-based population as the study sample. First, the channel catfish × blue catfish interspecific system provides a useful research system for the understanding of columnaris resistance because they exhibit clear contrast in their phenotypes (Arias et al. 2012). Segregation of genotypes in F2, along with the highly contrasted phenotypes, provides a good system for QTL analysis. Second, family-based association mapping is usually more powerful in detecting QTLs, since the lack of recombination between a QTL and linked marker increases the power of detection (Mackay and

Powell 2007). Third, family-based association mapping is preferred to detect rare markers related with QTLs (Mackay and Powell 2007).

1200 fish (average body weight 55.3 grams) were randomly obtained from Auburn University Fish Genetics Facility and acclimated for one week in the aerated flow-through water (240×60×45cm (L×W×H)). The average water temperature was 25℃. A total of 340 backcross progenies were selected from the extremes of the disease resistance distribution of the 1200 fish based on the selective genotyping method (Darvasi and Soller 1992). The first 169 fish that died of columnaris were continuously sampled as the susceptible group. When the fish lost balance, blood samples were collected. 171 fish that survived from the disease and showed no symptoms were selected randomly as the resistant group.

The bacteria challenge procedure was conducted as previously described (Sun et al. 2012). The bacteria *F. columnare* were provided by the Aquatic Microbiology Laboratory, Auburn University. To get a single *F. columnare* colony, several fish were experimentally infected with a virulent isolate (BGFS-27; genomovar II) (Olivares-Fuster and Arias 2011) and bacteria were re-isolated from one symptomatic fish after confirmed visually and biochemically. We selected BGFS-27 as representative of *F. columnare* for the present study, to which the hybrid was more resistant than blue and channel catfish (Arias et al. 2012). A single colony was cultured in modified Shieh broth for 24 h in a shaker incubator (100 rpm) at 28 ℃. The final concentration of the bacteria was determined using colony forming unit (CFU) per mL. Challenge experiments were then conducted by immersion exposure for 2h at a final concentration of 3×10^6 CFU/mL. Control fish were treated with identical procedures except that they were exposed to sterile modified Shieh broth.

3.3.3 DNA isolation, genotyping, and quality control

DNA was isolated from blood sample using standard protocols. After incubated at 55°C about 10h, the blood cells were broken by cell lysis solution first. Protease K and protein precipitation solution were used to remove the proteins. Next, DNA was precipitated by isopropanol and collected by brief centrifugation, washed twice with 70% ethanol, air-dried, and resuspended in TE buffer (pH 8.0). After quantified using spectroscopy by Nanodrop (Thermo Scientific) and checked by 1% agarose gel electrophoresis stained with ethidium bromide for integrity, DNA was diluted to 50 ng/uL.

We have developed a catfish 250K SNP array using Affymetrix Axiom genotyping technology (Liu et al. 2014; Liu et al. 2011). Genotyping using the catfish 250K SNP array was performed at GeneSeek (Lincoln, Nebraska, USA). The informative SNPs in this GWAS were distributed across the catfish genome at an average interval of 3.6 Kb. No sample was excluded due to low quality or low call rate (<95%). 214,797 SNPs were kept after filtering out SNPs with an inheritance or genotyping error, a minor allele frequency (MAF) <5%, or a call rate < 95%.

3.3.4 Statistical Analysis

Statistical analysis was carried out using the SVS software package (SNP & Variation Suite, Version 8.0). Pairwise linkage disequilibrium (LD) for the backcross progeny population was calculated according to r^2 value. LD pruning was conducted with a window size of 50 SNPs, a step of 5 SNPs, and r^2 threshold of 0.5, resulting in 14,420 independent SNP markers. The population structure was assessed by principal component analysis with the independent SNP markers. EMMAX (Efficient Mixed-Model Association eXpedited) analyses using all SNPs were conducted with the first two principal components and the fish body weight as covariates (Kang et al. 2010).

The threshold P-value for genome-wide significance was calculated based on Bonferroni-correction with the estimated number of independent markers and LD blocks (Duggal et al. 2008). A Manhattan plot of the P-value results was produced using the SVS software. Although channel catfish and blue catfish are two species, apparently their genome architecture is extremely similar according to our former studies and unpublished data (Ninwichian et al. 2012). Thus, the genetic marker map was constructed according to channel catfish genome sequence.

3.3.5 Sequence analysis

The upstream and downstream genes of the significant SNPs were determined. GENSCAN program(Burge and Karlin 1997) and FGENESH+ (Salamov and Solovyev 2000) were used to analyze the catfish genome sequences (unpublished data) that surround the SNPs to identify the upstream and downstream genes. The identified genes were annotated by searching against the non-redundant protein database (Altschul et al. 1990). Genomicus (Muffato et al. 2010) was utilized to construct the synteny of the counterpart genes from zebrafish to provide evidence for orthology.

3.4 Results

3.4.1 Mortality rate

The accumulative mortality rate was 45.7%. A total of 1,200 channel catfish from a mix of six families were pooled and communally challenged. The mortalities started 46 hours after challenge and peaked approximately 218 hours after challenge (Figure 1). Based on the selective genotyping method (Darvasi and Soller 1992), the blood samples of the first 169 dead fish were collected, serving as the “susceptible” fish. After 12 days of challenge, 652 fish survived. From

the survivors, 171 fish without symptoms of columnaris were randomly collected as the "resistant" fish.

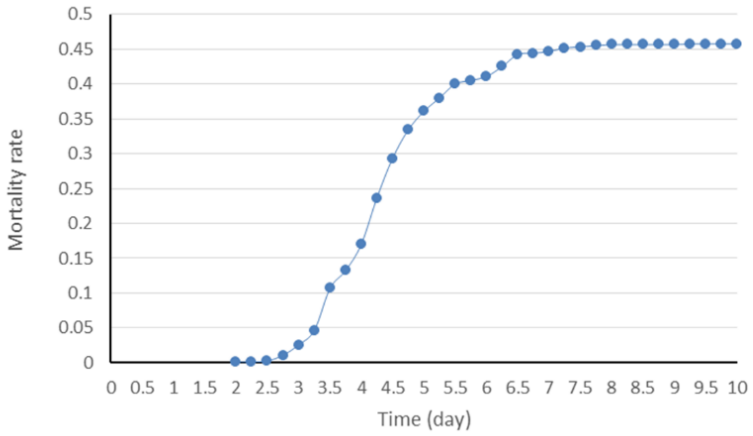

Figure 1 Mortality rate of hybrid catfish after *Flavobacterium columnare* infection.

3.4.2 Sample structure

The founders of the 6 families were known before the experiment, but the offspring were mixed for communal culture. In order to assign the genotyped fish to each of the six families, cluster analysis was conducted according to the IBS kinship matrix (Table 2). With known family pedigree, principal component analysis was conducted using PC scores of samples as coordinates to visualize the sample structure. As shown in Figure 2, apparently, families 4, 5, and 6 were highly related, while families 1, 2, and 3 were distantly related.

Table 2 The pedigree information of catfish samples used in this study.

Family ID	Dam	Sire	Sample number	Susceptible sample number	Resistant sample number
1	Channel 1	Hybrid 1	96	48	48
2	Channel 2	Hybrid 1	95	47	48

3	Channel 3	Hybrid 2	55	28	27
4	Channel 4	Hybrid 1	14	8	6
5	Channel 5	Hybrid 1	27	18	9
6	Channel 6	Hybrid 1	53	20	33

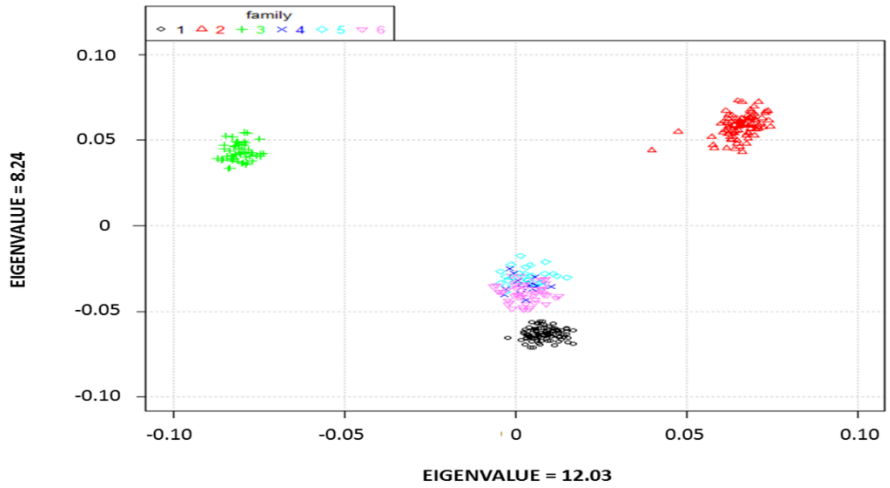

Figure 2 Sample structure identified by PCA with the first two principal components. The coordinates are the first two principal components.

3.4.3 Analysis of linkage disequilibrium (LD) blocks

The LD block was defined as a set of contiguous SNPs with the minimum pairwise r^2 value exceeding 0.50 (Gu et al. 2011). The number of independent SNP markers and LD blocks was 14,420. Thus the threshold P-value for genome-wide significance was 0.05/14420=3.47e-6 ($-\log_{10}$(P-value)=5.46). The threshold P-value for the significance of "suggestive association", which allows one false positive effect in a genome-wide test, was 1/14420=6.93e-5 ($-\log_{10}$(P-value)=4.16).

3.4.4 Linkage groups with associated QTLs for columnaris resistance

A Manhattan plot constructed using the marker positions and the corresponding $-\log_{10}$(P-value) is shown in Figure 3. Linkage group 7 harbors markers that are statistically significant at the genome level ($-\log_{10}$(P-value)>5.46). A second genomic region on linkage group 7 appears to

harbor suggestively associated markers, but is not statistically significant at the genome level. In addition to linkage group 7, linkage groups 12 and 14 appear to harbor SNP markers that are also suggestively associated with columnaris resistance, although not statistically significant at the genome level (Figure 3).

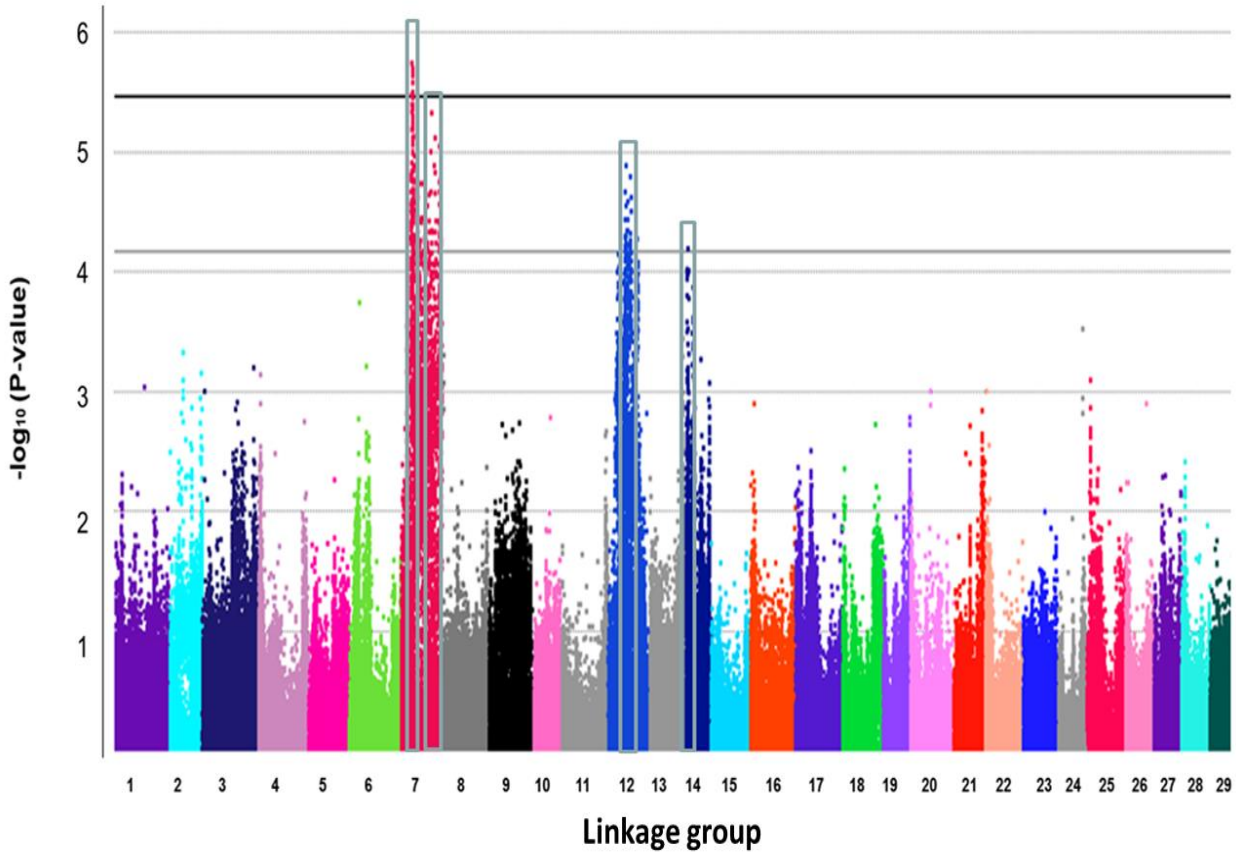

Figure 3 Manhattan plot of genome-wide association analysis for columnaris disease resistance. The black solid line indicates the threshold P-value for genome-wide significance. The gray solid line indicates the threshold P-value for the significance of "suggestive association". The four boxes indicate the associated regions.

3.4.5 Genomic region with significantly associated QTL for columnaris resistance

Additional analysis was conducted with the chromosomal region where the significantly associated QTL is located on linkage group 7. A set of 12 most significant SNPs are listed in

Table 3. These SNPs are all significantly associated with columnaris resistance at the genome level ($-\log_{10}$(P-value)>5.46). These SNPs are all located in a genomic region on linkage group 7 from 7,203,819 bp to 7,817,023 bp, spanning a total of approximately 620 Kb. Their minor allele frequencies vary between 0.24-0.39, and their nearby genes are listed in Table 3. Because family-based population with 8 founders were utilized, the haplotypes extend very long regions as expected (Figure 4).

Table 3 The significantly associated SNPs on linkage group 7.

SNP ID	Position (bp)	-log10(P-value)	Nearest gene
AX-86060479	7252290	5.75	upstream of *pik3r3b*
AX-85347098	7505396	5.70	*mast2* intron
AX-86056344	7633758	5.68	*adcyap1r1* exon
AX-85337705	7639440	5.65	*adcyap1r1* exon
AX-85377681	7569451	5.59	downstream of *akr1a1b*
AX-85265763	7503336	5.51	*mast2* intron
AX-85240370	7203819	5.48	upstream of *pik3r3b*
AX-85319066	7447662	5.47	*mast2* intron
AX-85346312	7618017	5.47	*adcyap1r1* intron
AX-85378689	7755734	5.47	*zc3h3* intron
AX-85233717	7803806	5.47	*zc3h3* intron
AX-85260766	7817023	5.47	*zc3h3* intron

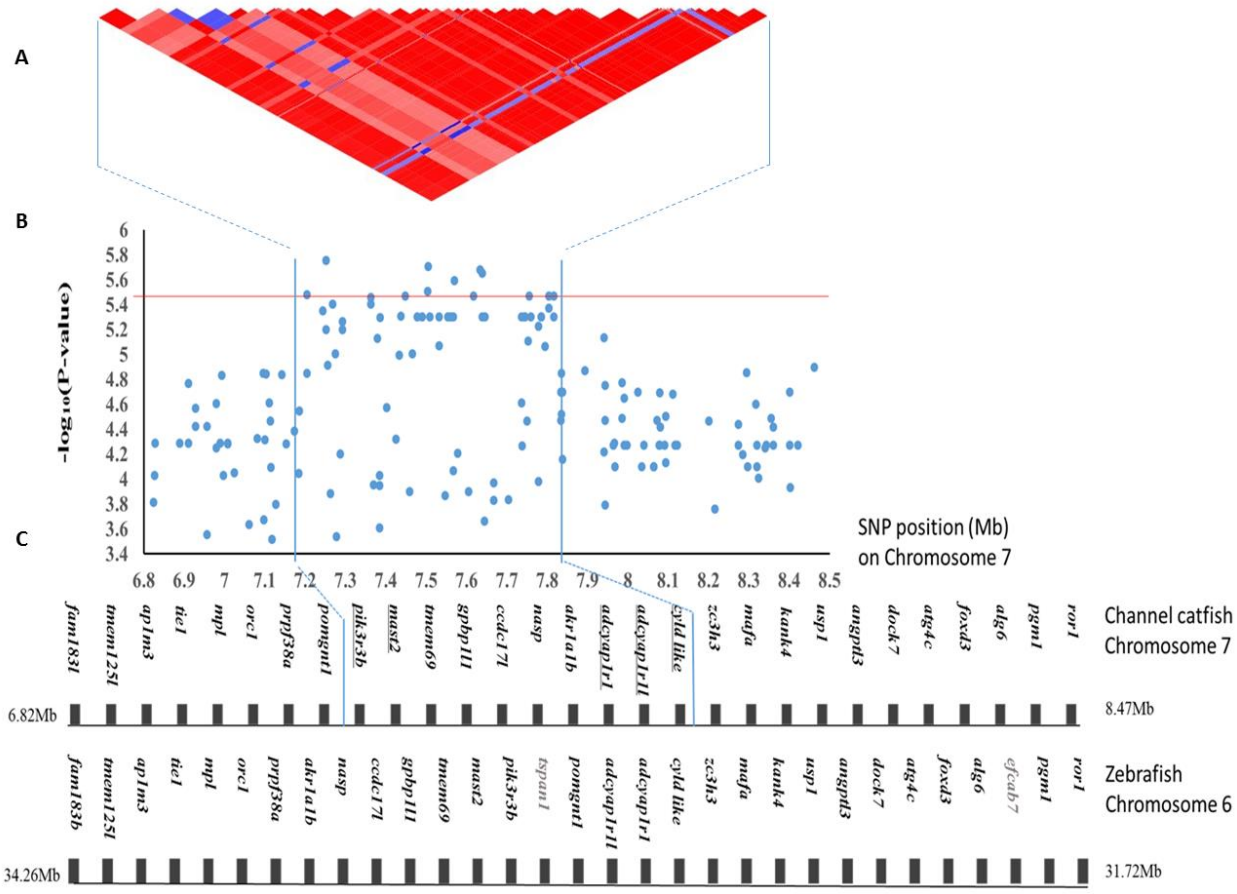

Figure 4 Regional genome scan for the QTL significantly associated with columnaris resistance on linkage group 7. A) Heat map of the LD between the most significant SNPs in the QTL region. B) Regional –log10(P-value) plot for the QTL. The horizontal red line indicates the threshold P-value for genome-wide significance. C) Synteny analysis comparing catfish and zebrafish. Candidate gene names are underlined. Genes with gray names are located in the region of zebrafish but not channel catfish.

EMMAX was used to investigate the contribution of the significantly associated QTL to the phenotype. Because of high correlation between SNPs in one locus (Figure 4), when analyzing the fraction of variance explained by the QTL, we chose only the most significant SNP (AX-86060479) to represent this region, which could explain 6.6% of the phenotypic variance. Nevertheless, binary phenotype and selective genotyping used in our study may cause potential overestimation of the QTL effect.

Based on the SNPs placed on the catfish 250K SNP panel (Liu et al. 2014), the parental origins of the SNPs could be determined. All the 12 significant SNPs are interspecific, which means on these loci, two species are simply fixed for alternate alleles. The resistant genotypes for the 12 SNP loci are homozygous with both alleles being originated from channel catfish.

3.4.6 Genes located within the significantly associated QTL region for columnaris resistance

The genes within the genomic region harboring the significant SNPs associated with columnaris resistance were annotated by BLAST analysis against the non-redundant protein database (Pruitt et al. 2007). To provide additional supporting evidence for the proper annotation of the region, syntenic analyses were also conducted to compare the gene contents in the genomic neighborhood around the significant SNPs. As shown in Figure 4, the conserved synteny was identified between the catfish and zebrafish genomes. The flanking genes of catfish are all conserved with zebrafish except *efcab7* and *tspan1* are missing in catfish and the gene orders are slightly different.

Within the 620 Kb region containing the most significant SNPs, a total of 10 genes were identified (Figure 4). Of the 10 genes, five genes were found to have known functions in immunity, and these include phosphatidylinositol 3-kinase regulatory subunit gamma b (*pik3r3b*), cylindromatosis-like (*cyld-like),* pituitary adenylate cyclase-activating polypeptide type 1 receptor (*adcyap1r1), adcyap1r1-like*, and microtubule-associated serine and threonine kinase 2 (*mast2*). In order to be sure all the candidate genes were included in the analysis, the extended genomic region was examined, and no gene was found with known function in immunity.

3.4.7 Suggestively associated QTLs

In addition to the significantly associated QTL on linkage group 7, three additional genomic regions were identified to contain SNPs suggestively associated with columnaris resistance ($-\log_{10}$(P-value)>4.16), although not statistically significant (Figure 3). As shown in Table 4, SNPs with relatively low P values were identified from the three suggestively associated regions on linkage groups 7, 12 and 14.

The first suggestively associated region is on linkage group 7. About 12 Mb downstream of the significantly associated region on linkage group 7, there is another locus suggestively associated with columnaris resistance. A series of SNPs on that region exhibit relatively low P values with the lowest ones listed on Table 4, with $-\log_{10}$ (P-value) ranging from 4.83-5.12. In addition to linkage group 7, there are SNPs on linkage group 12 that reach suggestive genome-wide significance (P-value<6.93e-5). The most significant SNP could explain 5.5% of the variance. Compared with the region strongly associated with the columnaris resistance on linkage group 7, the region detected on linkage group 12 is larger. According to the linkage map (Li et al. 2014), the recombinant frequency is very low on this region of linkage group 12. The distance of two most significant SNPs is 2.68 Mb on linkage group 12 expanding about 2 centimorgans, while in catfish, on average, 1 cM is equivalent to approximately 250 Kb. According to the suggestively associated interspecific SNPs on linkage group 12, the resistant genotypes of the SNPs are homozygous with two channel catfish alleles, like the SNPs on linkage group 7. On linkage group 14, there is only one SNP reaching suggestive Bonferroni genome-wide significance, explaining 4.6% of the variance. The candidate genes surrounding the most significant SNPs of these 3 loci are listed in Table 4.

Table 4 The most significant SNPs with the closest and candidate genes in 3 suggestively associated regions. The paralogs are marked by different symbols following the gene names.

Linkage group	SNP ID	SNP position (bp)	-log10 (P)	Gene name	Gene Position (bp)
7	AX-85432363	19715765	5.01	cysteinyl leukotriene receptor 2	19913410, 19914493
	AX-85417541	20399460	5.33	guanine nucleotide-binding protein (G protein) subunit beta 1*	19923349, 19927379
	AX-85231041	22082916	4.89		
	AX-85406722	22298185	4.83	voltage-dependent calcium channel subunit alpha 2/delta 2#	20083452, 20300502
	AX-85205344	22410724	5.12	hyaluronidase 2	20409454, 20413226
	AX-85278425	25230627	5.12	tumor suppressor candidate 2	20415941, 20421024
				diphosphoinositol polyphosphate phosphohydrolase 1	20432914, 20443526
				protein kinase C and casein kinase substrate in neurons protein 1	20469564, 20487008
				N-terminal EF-hand calcium-binding protein 1	22274849, 22324281
				alpha-1-syntrophin	22377042, 22408256
				probable G-protein coupled receptor 27	25065312, 25066872
12	AX-85394454	12067569	4.89	phosphatidylinositol 3-kinase regulatory subunit 5†	12246367, 12272207
	AX-85211547	14746472	4.79	phosphatidylinositol 3-kinase regulatory subunit 6†	12281638, 12300055
				guanine nucleotide-binding protein subunit beta 3*	12763269, 12767890
				voltage-dependent calcium channel gamma 6 subunit#	14286927, 14292214
				chondroitin sulfate proteoglycan 4	13668415, 13668975
14	AX-85234783	1601158	4.20	Spectrin beta chain, non-erythrocytic 2	1614033, 1619504

3.4.8 Correlation of the SNPs associated with columnaris resistance

Conditioned analyses were conducted to examine the correlation of the SNPs associated with columnaris resistance (Nishimura et al. 2012). Genotypes of the most significant SNP (AX-86060479) on linkage group 7 were included as a covariate in the mixed linear model. After conditioning, the $-\log_{10}$(P-value) of all the other SNPs on linkage group 7 dropped below 2.0, while the SNP P-values remained generally unchanged on the other linkage groups. On the linkage groups 12 and 14, there were also strong correlations among these associated SNPs within the same linkage group (data not shown).

3.5 Discussion

In this study, for the first time, we identified a significantly associated QTL on linkage group 7 and three additional suggestive QTLs on linkage groups 7, 12, and 14 for columnaris resistance. The significant QTL on linkage group 7 was narrowed down to a small region of 620 Kb. Therefore, in spite of being just an initial quantitative analysis for the complex disease resistance trait, this work is very significant because it has set the foundation for fine mapping of the columnaris resistance genes, providing basis for application of genome technologies for catfish aquaculture through marker- or genome-based selection. Additional fine mapping using larger or more families could narrow down the candidate genes underlining columnaris disease resistance.

A number of genes involved in the PI3K pathway were found to be within the significantly associated genomic region of 620 Kb, suggesting the involvement of PI3K pathway in the resistance against columnaris. Among the 10 genes found in the 620 Kb region, five were involved in PI3K gene pathway. These are *pik3r3b, cyld-like, adcyap1r1, adcyap1r1-like,* and

mast2. The phosphatidylinositol 3-kinase regulatory subunit gamma b (*pik3r3b*) is the closest gene to the most significantly associated SNP (AX-86060479) (Table 3). Although the causative SNP could be within any one of the 10 genes found in the 620 Kb region, the fact that PI3K pathway was reported to be involved in immunity and resistance makes them particularly interesting as candidate genes (Koyasu 2006). In addition, many genes found in the suggestive QTL regions were also involved in the PI3K pathway (see below), further increasing the likelihood of PI3K pathway involvement in the resistance against columnaris.

PI3 kinases have been known to play important roles in innate and adaptive immunity (Fukao et al. 2002; Jiang et al. 2000; Koyasu 2006; Vieira et al. 2001). For example, PI3K activity is important for NF-κB pathway activation in different mechanisms (Bone and Williams 2001; Reddy et al. 1997). It was also shown that the infectious agents can manipulate the PI3K pathway to create a favorable environment by various mechanisms (Ireton et al. 1999). PI3K is required for modifying the cytoskeleton dynamics, regulating membrane traffic, coordinating exocytic membrane insertion and pseudopod extension, which could be utilized by some pathogenic bacteria for entry into host cell (Cox et al. 1999; Ireton et al. 1996; Ireton et al. 1999; Lambotin et al. 2005; Pizarro-Cerda and Cossart 2006). Ireton et al.(1999) showed that the protein InlB from *Listeria monocytogenes* is an agonist of mammalian PI3K, which causes rapid increases in cellular amounts of $PI(3,4)P_2$ and $PI(3,4,5)P_3$. Lambotin et al. (2005) found *Neisseria meningitidis* requires cortactin recruitment by triggering a PI3K/Rac1 signaling to elicit an efficient invasion in non-phagocytic cells. Kierbel et al.(2005) reported *Pseudomonas aeruginosa* requires the PI3K and Akt pathway for internalization. It was reported that *Porphyromonas gingivalis* could activate PI3K, which blocks phagocytosis and promotes inflammation (Maekawa et al. 2014). In addition, it was shown that blockade or deficiency of

PI3Kγ significantly enhanced resistance against *Leishmania mexicana*, which revealed the unique role for Class I_B PI3K in parasite invasion (Cummings et al. 2012). The fact that *pik3r3b* is closest to the most significant SNP makes it an interesting candidate for future studies.

In addition to *pik3r3b* gene, four other genes (*cyld-like, adcyap1r1, adcyap1r1-like*, and *mast2*) in the PI3K pathway (Figure 5) found within the 620 Kb region containing the significant QTL have also been shown to play important roles in immunity. CYLD is a deubiquitylating enzyme that negatively regulates various signaling pathways by cleavaging lysine 63-linked polyubiquitin chains from several specific substrates (Massoumi 2010). For example, CYLD could regulate inflammation and the innate immune response via its inhibition of NF-κB activation (Sun 2010). Besides that, CYLD is a deubiquitinating enzyme for Akt and suppressed Akt activation (Yang et al. 2013). Gao et al.(2008) reported that CYLD also plays a role in the regulation of microtubule dynamics.

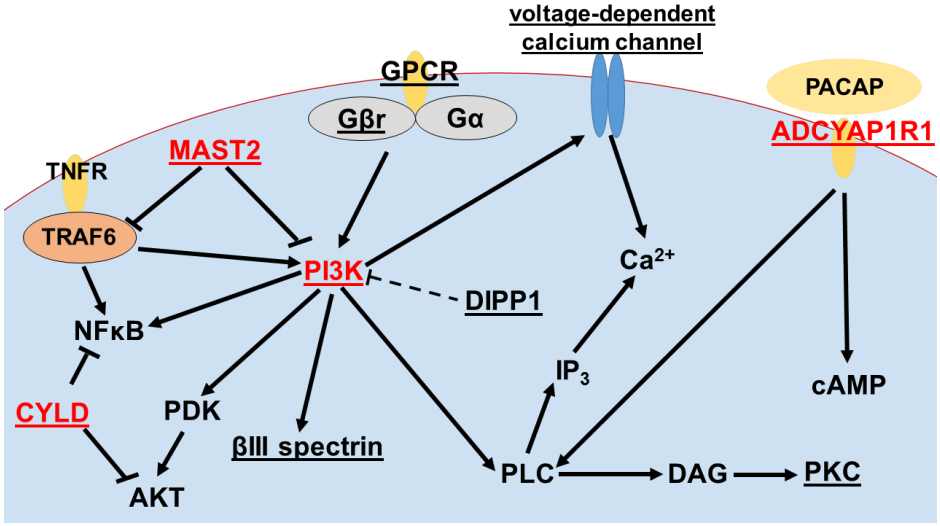

Figure 5 Signal transduction pathways involving PI3Ks and the other candidate genes. The candidate genes in the significant QTL are in red and underlined. The candidate genes in the suggestive QTLs are in black and underlined.

Pituitary adenylate cyclase-activating polypeptide type I receptor (ADCYAP1R1) is the pituitary adenylate cyclase-activating polypeptide (PACAP) specific receptor (Arimura and Shioda 1995). PACAP could activate adenylate cyclase and phospholipase C (PLC) through an interaction with ADCYAP1R1 and stimulates cAMP and inositol phosphate formation (Bodart et al. 1997; Romanelli et al. 1997). In fish larvae, Lugo et al.(2010) reported that the PACAP influences immune functions. They observed an elevated level of nitric oxide synthase-derived metabolites and total immunoglobulin M concentration in serum of juvenile catfish and tilapia after intraperitoneal injection of PACAP.

Microtubule-associated serine and threonine kinase 2 (MAST2) can interact with phosphatase and tensin homolog deleted on linkage group 10 (PTEN) which antagonizes PI3K-dependent signaling pathways (Downes et al. 2001; Terrien et al. 2012). MAST2 inhibits TNF receptor associated factor 6 (TRAF6) activity (Xiong et al. 2004), which represents a molecular bridge for innate immunity and adaptive immunity (Wu and Arron 2003). For example, TRAF6 could activate PI3K-dependent cytoskeletal changes and activate IκB kinase (IKK) in response to proinflammatory cytokines (Wang et al. 2006). Taken together, PI3 kinases themselves, or genes involved in PI3K pathway could play important roles in disease resistance.

Interestingly, many genes mapped within the suggestively associated QTL regions on linkage groups 7, 12, and 14 are functionally related to the genes mapped within the significant QTL region on linkage group 7, further supporting the possibility that PI3K pathway may be important for disease resistance against columnaris (Table 4, Figure 5). On linkage group 7, cysteinyl leukotriene receptor 2 is a G protein-coupled receptor (GPCR) with various functions such as modulation of chemokine gene transcription and calcium signaling (Sjöström et al. 2003; Thompson et al. 2008). G beta gamma activates the class I_B p110 gamma/p101 PI3K gamma on

the stimulation of GPCR (Brock et al. 2003). Voltage-dependent calcium channel subunit alpha-2/delta-2 gene was found in this region, and PI3K could enhance native voltage-dependent calcium channel currents (Viard et al. 2004). Diphosphoinositol polyphosphate phosphohydrolase 1 could cleave a beta-phosphate from the diphosphate groups in PP-$InsP_5$. PP-$InsP_5$ is similar to Ins(1,3,4,5)P_4, the headgroup of PI(3,4,5)P_3, which implied their competition relationship (Shears 2009). Chockalingam (1999) reported that alpha 1-syntrophin could bind PI(4,5)P_2. Hyaluronidase-2, a glycosylphosphatidylinositol-anchored receptor, could hydrolyze hyaluronic acid which could be degraded by *F. columnare* (Declercq et al. 2013). On linkage group 12, the genes within the suggestive QTL region seemed to be related with those within the QTL region on linkage group 7, because the paralogs of some genes in this region on linkage group 12 are found on linkage group 7 including phosphatidylinositol 3-kinase regulatory subunit 5 (*p101-PI3K*), phosphatidylinositol 3-kinase regulatory subunit 6 (*p87-PI3K*), guanine nucleotide-binding protein subunit beta 3, and voltage-dependent calcium channel gamma 6 subunit. On linkage group 14, spectrin beta chain, non-erythrocytic 2 *(βIII spectrin)* is located closest to the most significant SNP (AX-85234783). With a PH domain, βIII spectrin can bind PIP_2 and get involved in membrane skeleton (Holleran et al. 2001; Viel and Branton 1996). As presented in Figure 5, clearly many of these genes mapped in the significant QTL region or the suggestive QTL regions are involved in the related gene pathways.

It is notable that genes involved in the PI3K pathway were located together in “hubs” that were significantly associated with disease resistance. Theoretically, genes that are located together could be readily expressed in a coordinated fashion. However, here we do not have any evidence to indicate that the genes mapped within the QTLs involved in PI3K pathway are coordinately expressed. Analysis of RNA-seq data (Peatman et al. 2013) indicated that some

candidate genes indeed exhibited differences in baseline expression or after bacterial infection with columnaris. For instance, hyaluronidase-2 was expressed at a relative higher level in resistant fish than in susceptible fish before infection, and it was induced more in susceptible fish than resistant fish after infection. Guanine nucleotide-binding protein subunit beta-1, another gene that mapped within the suggestive QTL region, is expressed significantly higher in the susceptible fish than in the resistant fish. After infection, its expression in susceptible fish, but not in resistant fish, was drastically induced. However, because the experimental system is quite different, such expression data may not be directly transferable to our results here.

The most striking finding of our study was that the genes closest to the most significant SNPs were both positionally and functionally related, i.e., they are structurally organized as "functional hubs". Although it is unknown at present whether these genes are expressed in a coordinated fashion, it was reported that neighboring genes tend to have similar expression patterns and get involved in related functions (Michalak 2008; Sémon and Duret 2006; Williams and Bowles 2004). For instance, Schmid et al. (Schmid et al. 2005) elucidated that genes in close proximity are much more likely to be co-expressed than expected by chance. Future analysis for coordinated expression of genes involved in PI3K pathway is warranted.

The QTLs identified in this study explained a limited fraction of the phenotypic variance of columnaris disease resistance. Firstly, the population specificity of QTLs is the most important reason why our family-based association mapping cannot detect all the QTLs associated with columnaris resistance (Luo et al. 2013). Even within the same strain, various families showed drastically different susceptibilities to columnaris disease (LaFrentz et al. 2012). Secondly, segregating alleles within one species may lead to decreased power of analysis, especially in the case that one parental species systematically carries resistance alleles while the

other one carries susceptibility alleles (Ledur et al. 2009). Thus the region cannot be detected with strong significance using intraspecific SNPs. The associated SNPs on linkage group 14 are intraspecific, so the effect of the locus may be underestimated and we cannot infer the origins of the resistant alleles. Thirdly, because of the lack of recombination between nearby QTLs, the locus with minor effect cannot be detected if the favorite allele on a close QTL with a major effect have a different origin. Lastly, but not leastly, genome level variations such as allele variations can account for only a fraction of phenotypic variations. Gene expression regulations at various 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.

3.6 Conclusions

In summary, our GWAS using backcross interspecific hybrid population allowed mapping of associated QTLs and estimation of their effects for columnaris resistance. On linkage groups 7 and 12, the resistant genotypes were homozygous with both alleles from channel catfish. Examination of genes in the mapped QTL regions allowed further analysis of candidate disease resistance genes. It appears that signal transduction pathways involving PI3Ks may play a crucial role for disease resistance against columnaris. This notion is not only supported by the presence of PI3K pathway genes in the significantly associated QTL on linkage group 7, but also by the fact that many genes within the suggestive QTL on linkage group 12 are paralogs of those found on linkage group 7. In addition, many genes found within suggestive QTLs on linkage groups 7, 12 and 14 are also involved in PI3K pathway. Future studies are required for the identification of the causative genes for disease resistance. For example, GWAS using larger or

more families can be conducted to increase the power and resolution. Ultimately, gene knockout experiments are needed to demonstrate the candidate genes as the disease resistance genes.

The most interesting discovery of this work is that functionally related genes that may be responsible for columnaris disease resistance are located closely in a limited number of positions, forming “functional hubs”. Future analysis of expression of genes in the PI3K pathway in relation to the resistance phenotype should determine whether the co-localized and functionally related genes are indeed expressed in a coordinated fashion.

Chapter 4 Mapping of genetic regions for head size in catfish: the involved GTPase pathway is evolutionarily conserved for skull morphometric traits

4.1 Abstract

Skull morphology is fundamental to evolution and biological adaptation of species to its environments. With aquaculture fish species, head size is also important for economic reasons, because it has a direct impact on fillet yield. However, little is known about the underlying genetic basis. Catfish is the primary aquaculture species in the United States. In this study, we performed a genome-wide association study using the catfish 250k SNP array with backcross hybrid catfish to identify the QTLs for head size (head length, head width, and head depth).

Several QTLs were identified, including one significantly associated region on linkage group (LG) 9 and four suggestively associated regions on LGs 7, 16, 28 for head length, five significantly associated regions on LGs 5, 7, 9, 29 and two suggestively associated regions on LGs 7, 27 for head width, and one suggestively associated region on LG26 for head depth. Two of the QTLs on LG7 were associated with both head length and head width. Phenotypic variance explained by the associated SNPs for head length, head width, and head depth were 0.16, 0.18, and 0.02 from the associated regions respectively. It is notable that each of these associated regions is rich of small GTPase related genes. Comparative analysis indicated that small GTPase pathway genes are also involved in skull morphology in various other species ranging from amphibian to mammalian species, suggesting evolutionary conservation of small GTPase pathway in the control of skull morphologies.

4.2 Background

Skull morphology and body conformation are fundamental to evolution and biological adaptation of species to its environments. Species evolve to have different head shapes and sizes in response to their environments, and in relation to their behavior and mode of survival. As such, skull morphology and body conformation have been widely studied in various species. Wunnenberg-Stapleton et al. (1999) first reported the involvement of small GTPases RhoA and Rnd1 in control of head formation in *Xenopus*. Later, a number of studies were conducted in canine. As a companion species, dogs have been artificially bred and selected to have hundreds of breeds with various overall sizes and various head shapes and sizes, and their head sizes and shapes were found to be related to their behavior and personality (Schoenebeck and Ostrander 2013). The finding that selection of a single gene, *insulin-like factor I,* is largely responsible for the huge variations in shapes and sizes in dogs astonished many scientists (Sutter et al. 2007). Since then, great efforts were devoted to the analysis of head shapes and sizes in dogs in order to understand the genomic basis underlining the large difference in skull shapes and sizes (Schoenebeck and Ostrander 2013). GWAS allowed mapping of QTLs controlling head shapes in eight chromosomes in dogs. As the canine genome is available, further analysis of the QTL regions allowed identification of candidate gene *BMP3* for skull shapes (Schoenebeck et al. 2012). Recently, seven QTLs were identified for skull size and 30 QTLs were identified for skull shape in mouse (Maga et al. 2015). Analysis of the genes within these QTLs suggested that genes involved in the small GTPase pathway were among the "high priority" candidate genes, such as *Arhgap31*, *Fgfr3*, and *Kif7*.

With aquaculture fish species, analysis of head sizes is important not only for understanding of evolution and biological adaptation, but also for economic reasons. Head shapes and sizes influence fillet yield directly. Smaller head and uniform body conformation provide a greater proportion of fillet, and thus selection for smaller head and uniform body conformation is of great aquaculture value.

Genetic analysis of body shape has been conducted in fish species including common carp (Laghari et al. 2014), sea bass (Massault et al. 2010), and hybrid carp generated from silver carp x bighead carp (Wang et al. 2013a). However, limited by the number of markers, these findings are far from the requirements of marker-assisted selection (MAS), and little is known about genetic mechanisms for head shapes and sizes with aquaculture species. Channel catfish is the major aquaculture species in the United States. In recent years, hybrid catfish, produced by mating female channel catfish with male blue catfish, has become the breed of choice, because the F1 hybrid exhibits a number of superior traits due to heterosis including faster growth, enhanced disease resistance, and greater fillet yield (Argue et al. 2003; Dunham et al. 2008). Channel catfish in general has a relatively larger head than blue catfish. Therefore, the channel catfish X blue catfish hybrid system offers a great model to study head shapes and sizes. Understanding of genomic regions for head shapes and sizes in catfish will allow us to determine the level of evolutionary conservation in the controlling mechanisms. In addition, the linked markers will allow marker-assisted selection or marker-guided introgression. Traditional selective breeding has been used to enhance processing yield in catfish, but the progress has been limited due to low selection intensity and accuracy and low heritability (Argue et al. 2003).

GWAS has been regarded as a powerful strategy for the identification of markers associated with traits of interest with high resolution, but it has been rarely used in aquaculture

species (Geng et al. 2015). Recent development of a number of genomic resources has made such work feasible, including a large number of SNPs (Sun et al. 2014) and the 250K SNP arrays (Liu et al. 2014). In this study, we explored the genetic architecture for catfish head size using GWAS for the first time, and here we report the identified QTLs and their associated genes within the highly associated genomic regions.

4.3 Methods

4.3.1 Ethics statement

All experiments involving the handling and treatment of fish were approved by the Institutional Animal Care and Use Committee (IACUC) at Auburn University. Blood was collected after euthanasia. All animal procedures were carried out according to the Guide for the Care and Use of Laboratory Animals and the Animal Welfare Act in the United States.

4.3.2 Experimental fish and sample collection

The study population was the Auburn University one year old catfish generated from backcross of male F1 hybrid catfish (female channel catfish x male blue catfish) with female channel catfish. The population consisted of five families (Table 5). 386 fish in total (average body weight 53 grams ranging from 14g to 180g) were randomly obtained from Auburn University Fish Genetics Facility and blood samples were collected. The head size, including head length, head width, and head depth, was measured as the trait of interest (Figure 6). Head length is the horizontal distance between maxillary symphysis and posterior bony edge of operculum. Head width is the distance between the two sides of posterior bony edges of operculum. Head depth is the vertical distance from top to bottom of skull across posterior bony edge of operculum.

Table 5 The pedigree information of catfish samples used in this study.

Family ID	Dam	Sire	Sample number
1	Channel 1	Hybrid 1	96
2	Channel 2	Hybrid 1	91
3	Channel 3	Hybrid 2	55
4	Channel 4	Hybrid 1	51
5	Channel 5	Hybrid 1	93

Figure 6 Morphometric measurement of catfish skull.

4.3.3 DNA isolation, genotyping, and quality control

DNA was isolated from blood sample using standard protocols. After incubated at 55°C about 10h, the blood cells were broken by cell lysis solution first. Protease K and protein precipitation solution were used to remove the proteins. Next, DNA was precipitated by isopropanol and collected by brief centrifugation, washed twice with 70% ethanol, air-dried, and resuspended in TE buffer (pH 8.0). After quantified using spectroscopy by Nanodrop (Thermo Scientific) and checked by 1% agarose gel electrophoresis stained with ethidium bromide for integrity, DNA was diluted to 50 ng/uL.

A catfish 250K SNP array has been developed using Affymetrix Axiom genotyping technology with markers distributed across the catfish genome at an average interval of 3.6 Kb (Liu et al. 2014). Genotyping using the catfish 250K SNP array was performed at GeneSeek (Lincoln, Nebraska, USA). No sample was excluded due to low quality or low call rate (<95%). 190,706 SNPs were kept after filtering out SNPs with any genotyping error, a minor allele frequency (MAF) <5%, or a call rate < 95%.

4.3.4 Statistical analysis

Statistical analysis was carried out using the SVS software package (SNP & Variation Suite, Version 8.0) and PLINK (V*ersion 1.07)* (Purcell et al. 2007). The LD block was defined as a set of contiguous SNPs with the minimum pairwise r^2 value exceeding 0.50. In order to generate a set of independent SNPs, pairwise linkage disequilibrium for the backcross progeny population was calculated according to r^2 value. LD pruning was conducted with a window size of 50 SNPs, a step of 5 SNPs, and r^2 threshold of 0.5. Assuming each LD block represents one independent set of markers, the number of independent SNPs and LD blocks was 6044. To visualize the sample structure, principal component analysis with the independent SNP markers

was conducted, and the plots representing the sample structure were constructed with the first three principal components (Figure 7).

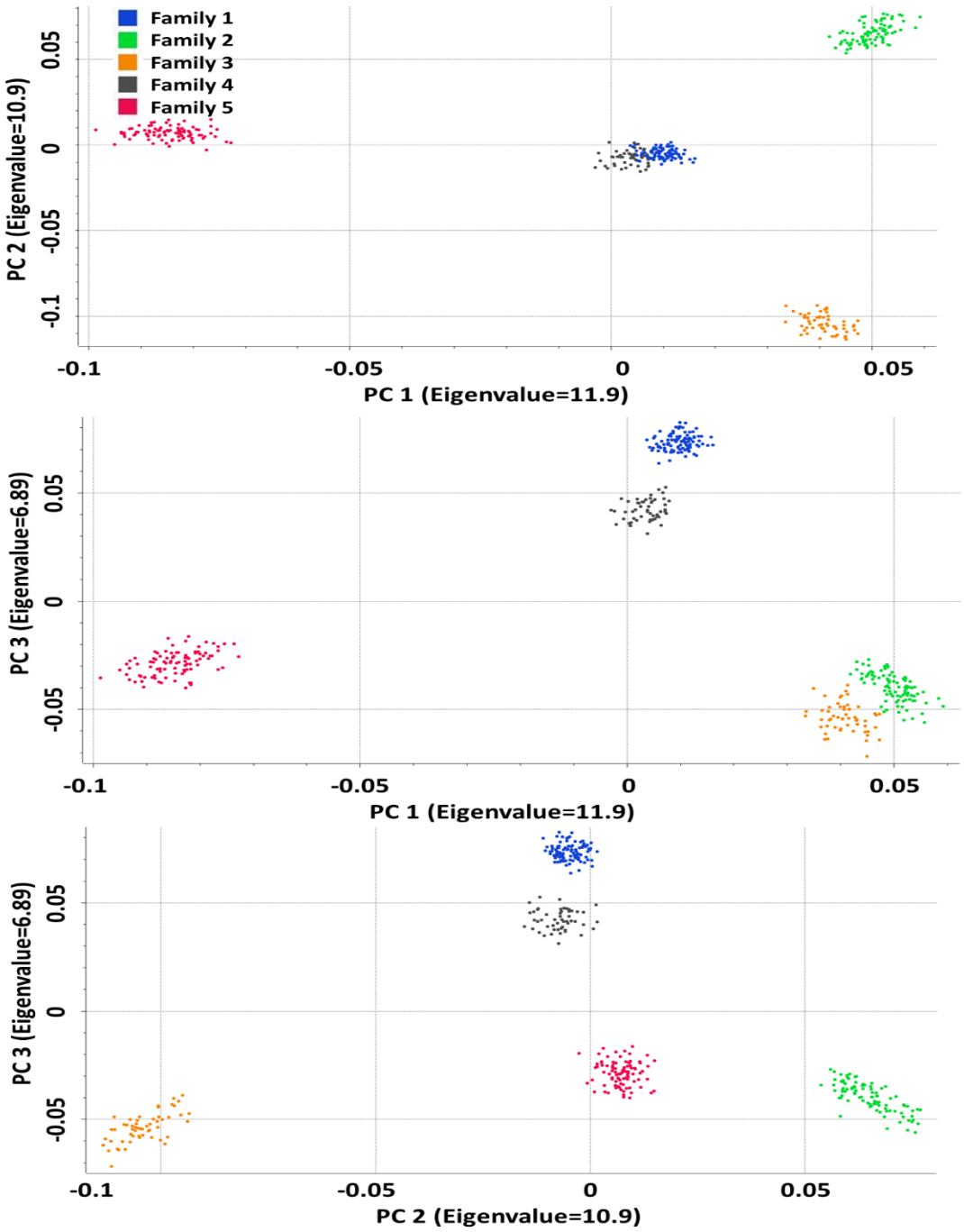

Figure 7 Sample structure identified by PCA with the first three principal components using sample genotypes.

A two-step GWAS procedure was performed. First, to eliminate the effect of body weight and family phenotypic stratification, the phenotypic data in the backcross population were adjusted with cubic root of body weight by simple linear regression within each family after outliers were filtered out. Second, the residuals were used as adjusted phenotypes to carry out genome-wide association analysis. Two methods were utilized to compare their performance in this step. The first method is EMMAX (Efficient Mixed-Model Association eXpedited) analyses (Kang et al. 2010). It was conducted in SVS without any covariates other than SNPs. The second method is family-based association test for quantitative traits (QFAM) conducted in PLINK (Abecasis et al. 2000; Fulker et al. 1999; Purcell et al. 2007).

A Manhattan plot was produced using the SVS software. The genetic marker map was constructed according to channel catfish genome sequence (version Coco1.0, Liu et al., in review), since genome architectures of channel catfish and blue catfish are extremely similar with same sets of parallel chromosomes according to our former studies and unpublished data (Kucuktas et al. 2009; Liu et al. 2003; Ninwichian et al. 2012). The significance level for genome-wide significance was set as 0.05/6044=8.27e-6 ($-\log_{10}$(P-value)=5.08) based on Bonferroni correction. The threshold of $-\log_{10}$(P-value) for suggestive association was arbitrarily set as 4.

4.4 Results

4.4.1 Phenotypes

Summary of original observations and adjusted phenotypes for head length, head width, and head depth was shown in Table 6. To cover individuals with a wide range of body size, samples were utilized with body weights varying from 14 g to 180 g. Because body weight could

explain over 70% variance for all the three traits, the effect of body weight should be eliminated before studying the association between markers and head sizes. The phenotypic data were adjusted with cubic root of body weight by simple linear regression. After adjustment, the means of the three traits were approximately 0. Adjusted head length ranged from -0.78 cm to 0.50 cm with standard deviation of 0.23 cm. Adjusted head width ranged from -0.74 cm to 0.43 cm with standard deviation of 0.17 cm. Adjusted head depth ranged from -0.61 cm to 0.77 cm with standard deviation of 0.20 cm. Adjusted phenotypes were utilized for further study.

Table 6 Summary of original observation and adjusted phenotype for three traits. N=386. SD, standard deviation; Min, minimum; Max, maximum.

	Original observation				**Adjusted phenotype**			
	Mean	SD	Min	Max	Mean	SD	Min	Max
Body weight /g	53.0	25.4	14	180	-	-	-	-
Head Length /cm	3.42	0.50	1.97	5.16	0	0.23	-0.78	0.50
Head Width /cm	2.53	0.41	1.38	3.75	0	0.17	-0.74	0.43
Head Depth /cm	2.34	0.40	1.38	3.75	0	0.20	-0.61	0.77

4.4.2 Determination of optimal model for analysis: EMMAX versus QFAM

The Manhattan plots generated from EMMAX and QFAM were shown in Figure 8, Figure 10, and Figure 11. Generally, the association results generated by EMMAX and QFAM were positively correlated, but QFAM provided more power than EMMAX. For example, LG9 harbored one significantly associated region for head length, and LG16 harbored one suggestively associated region according to EMMAX (Figure 8). The same regions were also identified by QFAM, but three more suggestively associated regions were identified using QFAM, including two on LG7 and one on LG28. The comparison of two methods for head width and head depth generated similar results. Therefore the QFAM performed better than EMMAX

with our family-based samples, considering QFAM had more power than EMMAX in detecting associated QTLs. In the following sections, we will describe the characters of identified regions according to the results generated from QFAM, unless otherwise noted.

4.4.3 Genetic regions associated with head size

For head length, significantly associated SNPs were identified around position 25.0 Mb of LG9 (Figure 8, Figure 9, and Table 7). The P-value of the most significant SNP reached 3.59e-6 ($-\log_{10}$(P-value)=5.45). In addition to LG9, three additional linkage groups, LG7, LG16, and LG28, were found to contain QTLs suggestively associated with head length (Figure 8 and Table 7). The associated regions of LGs 7 and 16 extended a long distance of over 1Mb. It may be caused by two or more candidate genes with a long interval that were located in the associated regions, just like the case shown in Table 7. Another reason may be the low recombination rate to break the linkage of nearby loci in the backcross progenies.

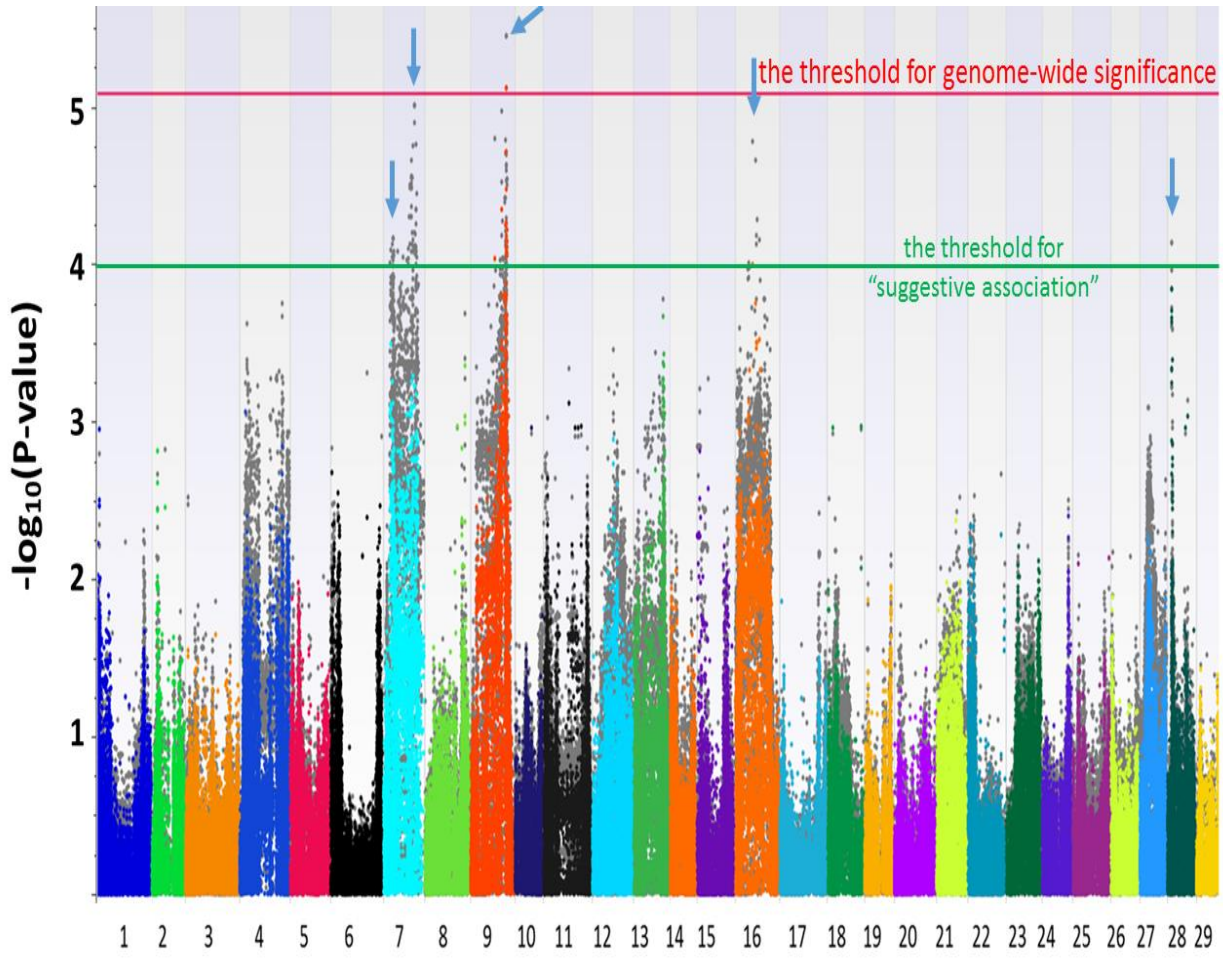

Figure 8 Manhattan plots for head length. The plots in different colors in the front layer were generated from EMMAX and the plots in gray in the back layer were generated from QFAM. The arrows indicate the associated regions.

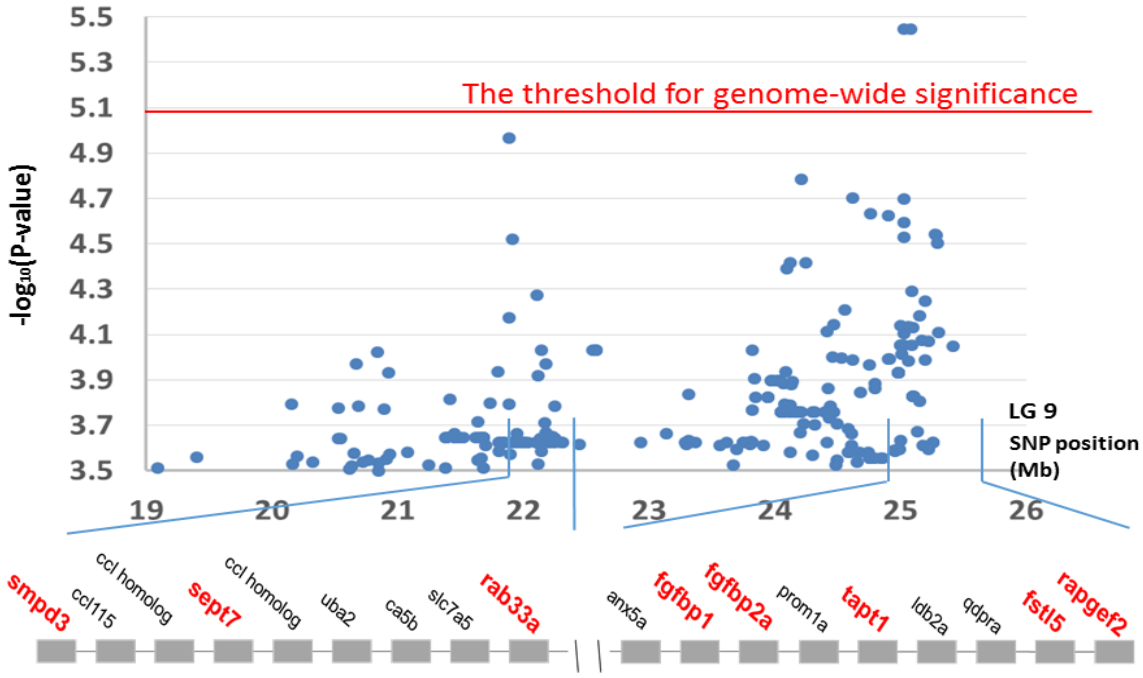

Figure 9 Regional genome scan for the QTL significantly associated with head length on LG 9.

Table 7 Information of regions associated with head length. "*" means the paralogs of the candidate genes were identified. "#" means the candidate genes are involved in small GTPases pathway.

LG	*SNP ID*	*SNP position*	*Beta*	*-log10 (P)*	*% variance*	*Gene position (Kb)*	*Gene Name*
9	85413092	25028079	0.13	5.45	5	21984-22010	**#sphingomyelin phosphodiesterase 3*
						22112-22134	*#septin 7*
						22301-22304	**#Ras-related protein rab-33a*
						24916-24918	**#fibroblast growth factor-binding protein 1*
						24925-24927	**#fibroblast growth factor-binding protein 2a*

						25001-25019	*transmembrane anterior posterior transformation 1*
						25226-25228	*follistatin-related protein 5*
						25600-25627	*#Rap guanine nucleotide exchange factor 2*
7	85385268	21742476	0.09	5.01	3	19874-19877	*#inhibin beta B chain like*
	85420069	22845333	0.07	4.76		19948-19951	*zinc finger protein GLI1*
	85308436	20362832	0.08	4.75		19974-19988	*zinc finger E-box-binding homeobox 2 like*
						20000-20035	**#Rho GTPase-activating protein 9*
						20133-20145	*#kinesin heavy chain isoform 5a*
						20217-20221	*transcription factor Sp7*
						20236-20249	*#ADP-ribosylation factor-related protein 1*
						20334-20377	**#Ral GTPase-activating protein subunit beta*
						20591-20593	*#alpha-1-syntrophin*
						21084-21109	*#cadherin 22 like*
						21633-21638	*#1,25-dihydroxyvitamin D(3) 24-hydroxylase, mitochondrial*
						21936-21939	*#regulator of G-protein signaling 9-binding protein*
						21997-22012	*synaptotagmin homolog*
						22240-22250	**#sphingomyelin phosphodiesterase 2*
						22522-22536	*diphosphoinositol polyphosphate phosphohydrolase 1 homolog*
						23035-23041	*#guanine nucleotide-binding protein G(I)/G(S)/G(T) beta-1b*
						23161-23168	*#guanine nucleotide-binding protein G(i) subunit alpha 2b*
						23547-	*#guanine nucleotide-*

						23552	*binding protein G(t) subunit alpha 1*
	85230603	6633722	4.03	4.16		4477-4715	*#protocadherin 9*
	85285134	5099134	-3.98	4.08		5920-5934	**#regulator of G-protein signaling 8*
						5946-5948	**#regulator of G-protein signaling 16*
						5975-5995	**#regulator of G-protein signaling 5*
						6015-6019	**#regulator of G-protein signaling 4*
						6299-6313	*discoidin domain-containing receptor 2a*
						6823-6841	*tyrosine-protein kinase receptor Tie 1*
						7188-7203	*#phosphatidylinositol 3-kinase regulatory subunit gamma b*
16	85254932	11508537	4.36	4.77	4	11590-11625	*# asap2*
	85252501	14167899	-4.30	4.66		11630-11634	*#integrin beta-1-binding protein 1*
	85340383	16798283	4.03	4.16		14152-14180	*#kinectin*
						16738-16750	**#sorting nexin 6*
						16755-16757	*cofilin 2*
28	86019627	2966557	4.01	4.13	4	2321-2324	*#Rho-related GTP-binding protein RhoV*
						2426-2429	**#regulator of G-protein signaling 6*
						2529-2551	*#signal-induced proliferation-associated 1-like protein 1*
						2602-2609	**#Ras-related protein Rab 15*
						2614-2627	*#farnesyltransferase subunit beta*
						2705-2717	*#G-protein coupled receptor 176*
						2978-3019	*nesprin 1*
						3120-3131	*#nuclear receptor-binding protein 1*

	3170-3174	*#protein Churchill*
	3265-3271	*ectonucleotide pyrophosphatase/phosphodiesterase family 7l*
	3287-3295	*matrilin 3*
	3462-3493	**#sorting nexin 9*

For head width, two associated regions around position 7Mb and position 23Mb on LG7 were identified, which were also associated with head length, implying the correlation between these two traits (Figure 8 and Figure 10). Apart from LG7, significantly associated regions were identified on LGs 5, 9, and 29. One more region on LG27 was suggestively associated with head width (Table 8).

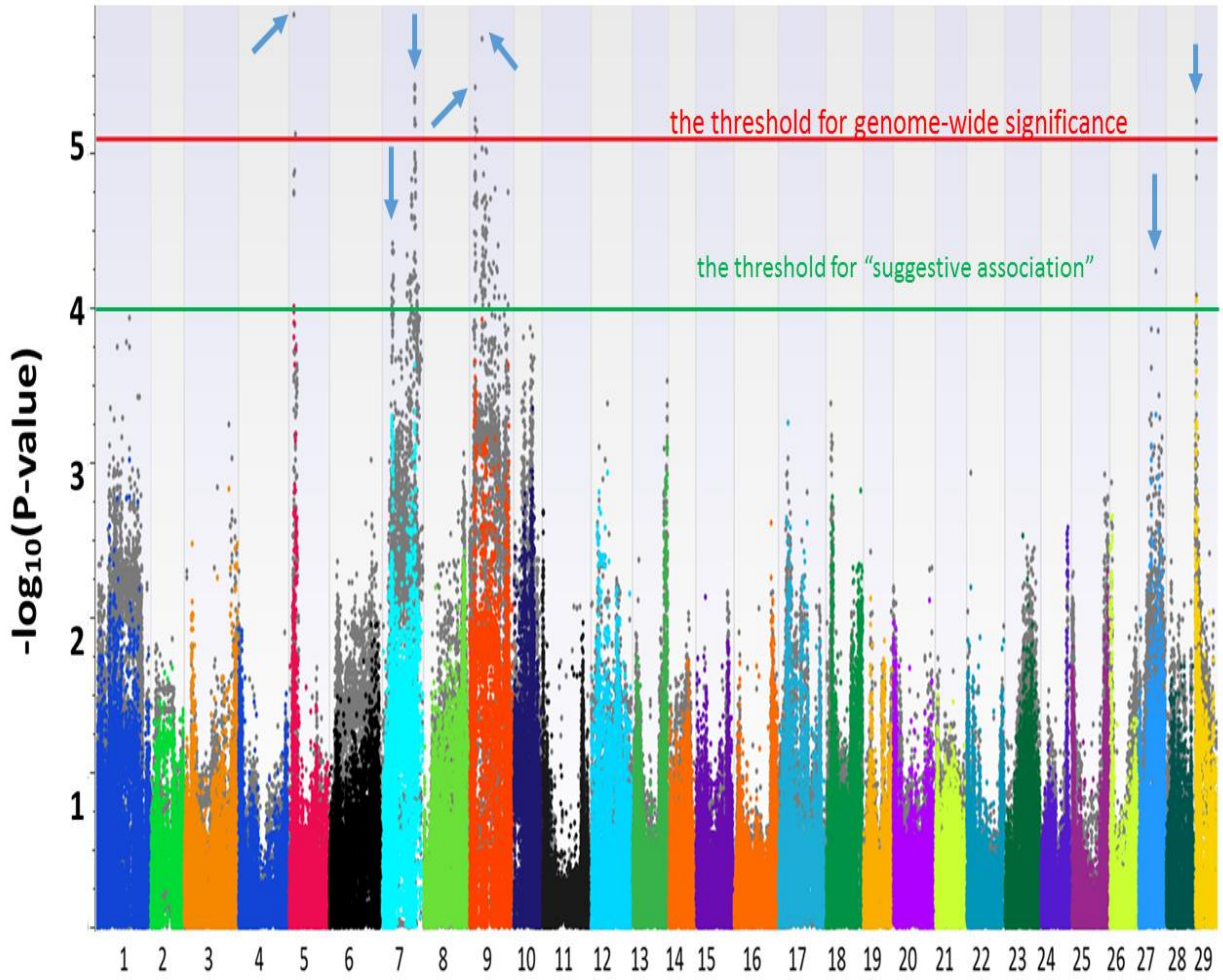

Figure 10 Manhattan plots for head width. The plots in different colors in the front layer were generated from EMMAX and the plots in gray in the back layer were generated from QFAM. The arrows indicate the associated regions.

Table 8 Information of regions associated with head width. Regions on Linkage group 7 explaining 0.03 of phenotypic variance for head width are not shown here, which are also associated with head length. "†" means the paralogs of the candidate genes were identified in the regions associated with head length. "#" means the candidate genes are small GTPases or related with small GTPases in function.

LG	*SNP ID*	*SNP position*	*Beta*	*-log10 (P)*	*% variance*	*Gene position (Kb)*	*Gene Name*
5	85206630	3471449	-0.10	5.89	4	3913-3940	*#integrin beta 3b*
						3942-3956	*synaptopodin 2-like b*

						4203-4208	*galanin receptor type 2*
9	85362293	8919023	0.09	5.73	4	8993-9042	*†#sorting nexin 25*
	85421035	3840292	0.07	5.42		3324-3336	*#fibroblast growth factor 16*
						3402-3429	*plastin 3*
						4028-4032	*regulator of cell cycle RGCC-like*
						4034-4038	*†#protocadherin 20*
						4058-4061	*†#Ras-related protein Rab 9b*
29	85361890	719786	0.09	5.20	4	539-542	*#fibroblast growth factor 20*
						607-613	*†#protocadherin 18*
						1049-1067	*†#Rho GTPase-activating protein 7*
27	85263029	12235962	0.09	4.23	3	11724-11737	*†#cadherin 15*
						12322-12326	*zinc finger protein SNAI2*
						12333-12343	*†#cadherin 1*

For head depth, although no genome-wide significant SNP was identified, one SNP (AX-85206421) with $-\log_{10}$(P-value) of 4.32 was located on LG26 (Figure 11, Table 9).

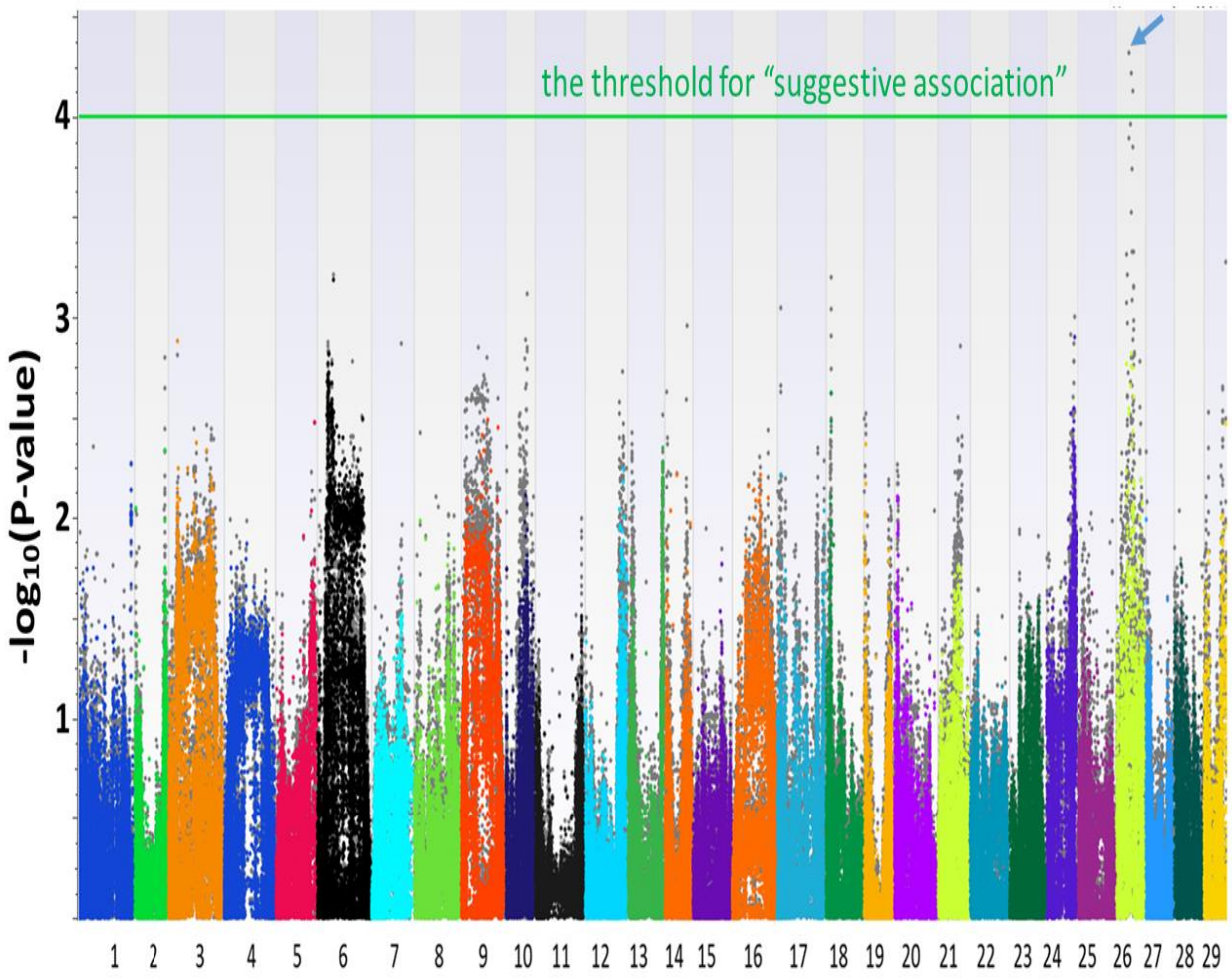

Figure 11 Manhattan plots for head depth. The plots in different colors in the front layer were generated from EMMAX and the plots in gray in the back layer were generated from QFAM. The arrows indicate the associated regions.

Table 9 Information of regions associated with head depth.

LG	*SNP ID*	*SNP position (Kb)*	*Beta*	*-log10 (P)*	*% variance*	*Gene position (Kb)*	*Gene Name*
26	85206421	8376	0.07	4.32	2	8270-8285	*Rho-related BTB domain-containing protein 2*
						8471-8503	*Ras GTPase-activating-like protein IQGAP2*
						9182-9206	*Ras GTPase-activating protein 1*

4.4.4 Candidate genes for head size

The regions surrounding the identified associated SNPs were examined for candidate genes according to their locations and functions. The genes within the genomic regions were predicted using FGENESH (Solovyev et al. 2006) and annotated by BLAST analysis against the non-redundant protein database (Altschul et al. 1990; Pruitt et al. 2007). Within the region significantly associated with head length, eight candidate genes related with bone development or head formation were identified (Table 7). Interestingly, like the candidate genes on the significantly associated region, most of the candidate genes on the suggestively associated regions are also functionally related with G protein, especially the small GTPases. Moreover, paralogs of the candidate genes were identified, including genes coding for small GTPase, small GTPase-activating protein (GAP), sphingomyelin phosphodiesterase, fibroblast growth factor-binding protein, regulator of G-protein signaling, and sorting nexin (Table 7).

In addition to LG7, the candidate genes for head width on LGs 5, 9, 27, and 29 are functionally related with those for head length, most of which are also involved in small GTPase pathway (Table 8). Some paralogs of candidate genes for head length were identified in the regions associated with head width as well, including *protocadherin-20 and -18, Ras-related protein Rab-9b, sorting nexin-25, Rho GTPase-activating protein 7,* and *cadherin-15* and *-1*. Some other genes like *integrin beta-3b,* and *fibroblast growth factor-16* and *-20* were also included in the associated regions functionally related with small GTPases.

For head depth, three candidate genes were identified including *Rho-related BTB domain-containing protein 2, Ras GTPase-activating-like protein IQGAP2,* and *Ras GTPase-activating protein 1,* which are all involved in small GTPase pathway

4.4.5 Phenotypic variance explained by associated SNPs

The phenotypic variance explained by associated SNPs of head size was estimated by EMMAX. Because of high correlation among SNPs on the same linkage group, when analyzing the fraction of phenotypic variance explained by the QTL, only the most significant SNP on the same linkage group was chosen. Thus the fraction of phenotypic variance explained may be underestimated. The fraction of variance of head length could be explained by the most significant SNP (AX-85413092) on LG9 is 0.05. In addition, the other suggestively associated regions could explain 0.11 in total (Table 7). The proportion of explained head width variance was 0.18 in total from significantly associated QTLs and suggestively associated QTLs (Table 8). Since there was only one suggestively associated QTL on LG26 for head depth, the proportion is as low as 0.02 from this QTL.

4.4.6 Conditioned analysis results

Conditioned analyses were conducted to examine the correlation of the SNPs associated with head size. The association test was conducted with the most significant SNPs associated with head size on each linkage group as a covariate (one SNP at a time). Because of the lack of recombination among SNPs on the same linkage group in the backcross population, the $-\log_{10}$(P-value) of SNPs on the same linkage group with the SNP included as covariate dropped drastically after conditioning, implying the SNPs on the same linkage group were highly correlated. For example, after the most significant SNP for head length on LG9 (ID AX-85413092) was included in the association test, the $-\log_{10}$(P-value) of SNPs on LG9 all dropped below 2, while the $-\log_{10}$(P-value) of SNPs on the other linkage groups generally did not change. Similar results were obtained for other associated SNPs. The independence of SNPs on different

linkage groups proved that no associated QTL was identified on a wrong linkage group caused by incorrect scaffolding in the genome sequence (Liu et al., in review).

4.5 Discussion

Head shape is not only important for evolutionary adaptation, but also highly relevant to aquaculture, as smaller head would translate into greater percentage of fillet yield. Therefore, selection of catfish with uniform body shape and smaller head is of significance for aquaculture production and profit margins. In this work, we used the high density 250K catfish SNP arrays and the backcross progenies for mapping the QTLs controlling head size. Five QTLs, one genome-wide significant and four suggestive, for head length were mapped to four linkage groups, LG7 (2 QTLs), LG9, LG16, and LG28. Seven QTLs for head width, five genome-wide significant and two suggestive, were mapped to five linkage groups, LG5, LG7 (2 QTLs), LG9 (2 QTLs), LG27, and LG29. One suggestive QTL for head depth was mapped to LG26. Interestingly, the two QTLs on LG7 were associated with both head length and head width, suggesting that they are pleiotropic.

The analysis of the functions of candidate genes reveals the crucial roles of small GTPase pathway in the control of head size. It is apparent that genes involved in small GTPase pathway were included in QTLs for head length, head width, as well as head depth. Although these QTLs were located on eight distinct linkage groups, genes involved in small GTPase pathway were the commonality found in all eight linkage groups (Table 7, Table 8, and Table 9).

Small GTPases were reported to be involved in bone morphogenesis with a variety of functions, including the dynamics of the actin cytoskeleton, cell adhesion, and membrane trafficking (Itzstein et al. 2011; Wunnenberg-Stapleton et al. 1999). They are involved in the

functions of some cell adhesion molecules, including cadherins, protocadherins and integrins (Watanabe et al. 2009). In our study, many candidate genes were involved in small GTPase pathway. For example, within the region associated with head length on LG9, eight candidate genes were found with functions related with head formation (Table 7 and Figure 9). One small GTPase gene and five genes with known functions highly related with small GTPase were identified in the region on LG 9, including *ras-related protein rab-33a* (*rab33a*), *rap guanine nucleotide exchange factor 2* (*rapgef2*) (Quilliam et al. 2002), *fibroblast growth factor-binding protein-1* (*fgfbp1*) and *-2a* (*fgfbp2a*) (Szebenyi and Fallon 1998), *sphingomyelin phosphodiesterase 3* (*smpd3*) (Aubin et al. 2005; Tomiuk et al. 1998), and *septin-7* (*sept7*) (Longtine and Bi 2003) (Figure 12). Apart from these genes, the other candidate genes within the associated region are involved in the bone morphogenetic protein (BMP) pathway, which is related with small GTPase pathway. Follistatin-related protein (FSTL) could bind actin and BMP, which is important in cartilage and bone development (Sidis et al. 2002). Transmembrane anterior posterior transformation 1 (Tapt1), regulated by BMP (McPherron et al. 1999), is speculated to be related with axial skeletal patterning during development (Howell et al. 2007). Moreover, on the other linkage groups associated with head size, most of candidate genes are known related with small GTPases, further proving the importance of small GTPase pathway to head formation.

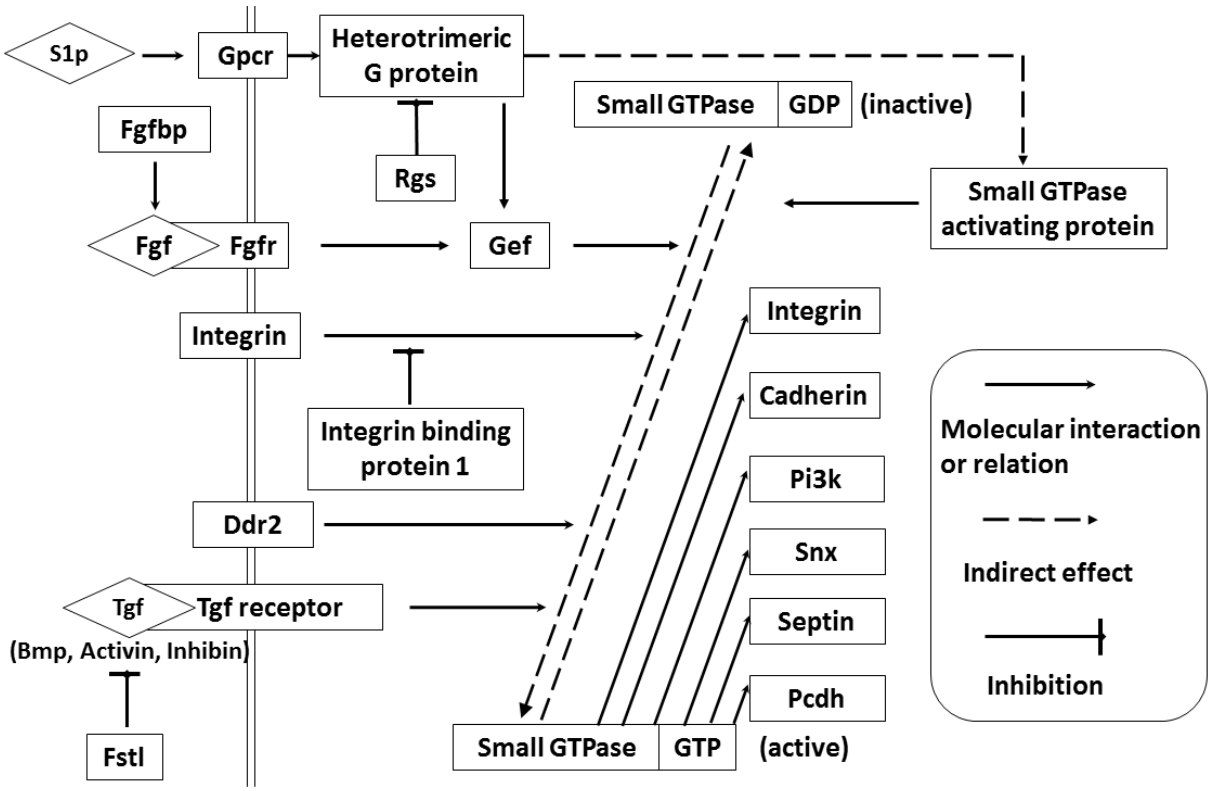

Figure 12 Signal transduction pathways involving small GTPases and the other candidate genes.

It is also notable that apart from G protein genes, some paralogs were identified in the regions associated with head size on different linkage groups, including gene family members of *small GTPase-activating protein*, *regulator of G-protein signaling*, *sphingomyelin phosphodiesterase, sorting nexin*, *protocadherin,* and *cadherin*.

Genes of the small GTPase pathway were previously reported to control skull shape and size in fish, frog, dog, mouse, and human. For instance, *bmp4* was reported to play an important role in coordinating shape differences in the cichlid fish oral jaw apparatus (Albertson et al. 2003). It was reported that small GTPases were important to cell adhesion and head formation in early *Xenopus* development (Wunnenberg-Stapleton et al. 1999). In human beings, mutation affecting *FGFR*, *RAB,* and *TGFBR* were associated with defects within the developing skull (Schoenebeck and Ostrander 2013). In dogs, eight QTLs were reported to be associated with

skull diversity (Schoenebeck et al. 2012). Schonenebeck et al. demonstrated that *BMP3* contained a likely causal variant. In addition to *BMP3*, we searched the other associated regions for candidate genes surrounding the proposed significant SNPs (Schoenebeck et al. 2012). In doing so, many candidate genes involved in small GTPase pathway were identified, coding for proteins including small GTPase, Ras guanine nucleotide exchange factor 1B (RASGEF1B), alpha-1-syntrophin (SNTA1), integrin alpha 11 (ITGA11), fibroblast growth factor 5 (FGF5), G-protein coupled receptor (GPR), kinesin-like protein (KIF), and insulin-like growth factor 1 (IGF1) (Figure 13). In mouse, *Itga2*, *Arhgap31*, *Gnai3*, *Fgfr3*, *Chd7*, and *Kif7* were identified within the QTLs linked with mouse skull shape (Maga et al. 2015). Close to *Gnai3*, *Itga2*, and *Chd7*, we also found *Gnat2*, *Itga1*, *Fst* and *Rab2a* according to mouse genome sequence, which are involved in small GTPase pathway (Figure 13). The fact that the genes involved in the small GTPase pathway were found within the QTL regions in species ranging from fish, frog, mouse, dog, and human suggested that the involvement of small GTPase pathway in control of head shape and size is evolutionarily conserved.

Figure 13 Regional scan of QTLs associated with head shape identified in mouse and dog (Maga et al. 2015; Schoenebeck et al. 2012). The homologs of mouse and dog candidate genes within the associated QTLs in catfish were also shown. Homologs were marked in the same color. Solid gray boxes indicate candidate genes. Dash lines under the boxes indicate several genes located in the interval are not shown.

The observation that many genes involved in the small GTPase pathways were found within QTLs in various species triggered us to determine if the genes were evolutionarily conserved to explain head shape variance in different species. By examining the genomic regions

associated with head shape in catfish, dog, and mouse, a total of 10 gene families involved in the small GTPase pathway were identified with candidate homologous genes in three species (Maga et al. 2015; Schoenebeck et al. 2012)(Figure 5) . It is clear that, with the exception of alpha-1-syntrophin, all the identified genes were not orthologous. For instance, small GTPases were mapped in the QTLs in two chromosomes of dogs, five linkage groups of catfish, and one chromosome of mouse. However, the genes within the dog QTLs were *RASSF3* and *CDC42*; the genes within the mouse QTL was *Rab2a*; and the genes within the catfish QTLs were *rab33a, rab9b, rab15, rhov,* and *rhobtb2*. Similarly, the situations were also true for *small GTPase activating proteins, follistatin, guanine nucleotide exchange factor, integrin, fibroblast growth factor, heterotrimeric G protein, G-protein coupled receptor,* and *kinesin*. Therefore, it appears that it is not orthologous genes that were identified to explain the variance of head shape in different species. The first reason for that may be that different landmarks were utilized to describe the head shape in three studies. Secondly, maybe not all the orthologous genes contain variants that could affect the phenotype in the sampled population within three species, despite the possibility that the orthologous genes are involved in head shape in three species, so the related regions cannot be mapped by GWAS. Thirdly, it is not guaranteed that all involved QTLs could be identified by GWAS.

We previously proposed the "functional hubs" within QTLs of columnaris resistance (Geng et al. 2015), where related genes in the same or similar pathways are physically together. Here once again, strong clusters of genes involved in the same pathway were also observed for head shape (Table 7, Table 8, and Table 9) (Geng et al. 2015; Michalak 2008). Although it is possible that just one causal gene is involved in each of the associated regions, it is also possible,

and even likely that the candidate genes in the functional hubs work together to regulate the involved traits, in this case, the head shape.

Family-based population is suitable for GWAS of most aquaculture species. In our study, samples based on five families were used. Usually, family-based design needs additional genotype information of parents, but catfish have high fecundities with thousands of progenies per spawn, which reduces the efficiency penalty for genotyping parents. Compared with unrelated sample, family-based population takes some advantages in identifying QTLs by GWAS. Firstly, the lack of recombination between QTLs and associated markers increases the power for detection (Mackay and Powell 2007). However, the tradeoff is that mapping resolution is reduced, which results in the long-extending regions of QTLs. To narrow down the regions, saturated SNP markers at a high density could be selected from local regions around the identified QTLs. Instead of genotyping SNPs on whole genome, genotyping local SNPs costs less, which allows larger number of samples to be included to detect rare recombination. The large number of offspring per spawning in most aquaculture species could ensure enough samples at a low cost. Thus fine mapping could be achieved by the two-step methods cost-efficiently. In order to improve brood stocks in catfish production by marker-assisted selection, further analysis on local SNPs is required to provide more accurate QTL information based on our preliminary data. Secondly, the clear pedigree information of family-based population design makes it much easier to control the confounding factor caused by population stratification. Moreover, family-based design obtains more power to detect rare variants, which only exist in specific families. Nevertheless, the pitfalls of family-based population should not be ignored. The family or population specification of QTLs is one of major reasons for the variance of

phenotype. The limited number of founders in the family-based samples may reduce power to detect QTLs.

The high fecundity makes family-based samples feasible and efficient for GWAS in most aquaculture species. However, for samples consisting of large families in aquaculture, the performances of commonly used test methods have not been compared. In our study, two methods, EMMAX and QFAM, were evaluated for family-based samples. EMMAX and QFAM are both effective in correcting population stratification. Population stratification is the major confounding factor causing false positive results. If the population stratification is not corrected, false positive results that are associated with population structure rather than the trait of interest could be detected as the associated markers falsely. Different from the universal applicability of EMMAX, QFAM is just applicable for family-based population to control population stratification. QFAM partitions the genotypes into between- and within- family components (Abecasis et al. 2000; Fulker et al. 1999). The within-family components could control stratification, and the true association results could be identified without the effect from stratification. Unlike QFAM, EMMAX calculates a pairwise relatedness matrix according to high-density markers to represent the sample structure at first. Then EMMAX could estimate the contribution of the sample structure to the phenotype, and detect associations without confounding effect generated from sample structure (Kang et al. 2010). EMMAX has been proven widely applicable for correcting family structure, as well as population structure and cryptic relatedness (Price et al. 2010). However, it has been shown that inclusion of the candidate markers to calculate the pairwise relatedness matrix could lead to loss in power because of double-fitting of the candidate markers in the model (Yang et al. 2014). In our study, EMMAX has less power compared with FBAT. Because of loss in power in EMMAX, we conclude that

QFAM performs better than EMMAX for our samples consisting of large families according to our results.

Our long-term goal is to enhance catfish stocks with favorite phenotype for head shape, incorporate such trait along with other traits such as disease resistance, and finally support a sustainable and profitable aquaculture industry. Parental fish with homozygous favorite alleles will be screened, making them immediately applicable to the catfish industry. To reach this long-term goal, genetic basis underling the traits must be understood, especially the accurate location of QTLs controlling the traits of interest. Detailed QTL information will then be used to improve brood stocks by marker-assisted selection, or introgression of valuable disease resistance QTLs from both channel catfish and blue catfish. In our study, GWAS could locate part of associated QTLs into several Mb, so fine mapping of QTL is still needed for a more efficient marker-assisted selection for these QTLs. Considering the high heritability, the improvement will be significant for the catfish industry.

4.6 Conclusion

This study investigated the genetic basis of head size of catfish backcross fingerlings. For head length, one genome-wide significant QTL was identified on LG 9. Besides LG 9, several suggestively associated QTLs were located on LGs 7, 16, and 28. For head width, significant SNPs were located on LGs 5, 7, 9 and 29, and suggestively associated SNPs were located on LGs 7 and 27. For head depth, only one suggestively associated region was identified on LG 26. Each of these associated regions is rich of small GTPase related genes, implying the crucial role of small GTPase pathway in controlling head size.

Note: Chapter 2 will be published in the book named as *Bioinformatics in Aquaculture*; Chapter 3 was published on BMC Genomics (Geng et al. 2015); Chapter 4 was submitted to *G3.*

References

Abecasis G, Cardon L, Cookson W (2000) A general test of association for quantitative traits in nuclear families. The American Journal of Human Genetics 66:279-292.

Albertson RC, Streelman JT, Kocher TD (2003) Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proceedings of the National Academy of Sciences 100:5252-5257.

Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ (1990) Basic local alignment search tool. Journal of molecular biology 215:403-410.

Argue BJ, Liu Z, Dunham RA (2003) Dress-out and fillet yields of channel catfish, Ictalurus punctatus, blue catfish, Ictalurus furcatus, and their F 1, F 2 and backcross hybrids. Aquaculture 228:81-90.

Arias CR, Cai W, Peatman E, Bullard SA (2012) Catfish hybrid *Ictalurus punctatus* x *I. furcatus* exhibits higher resistance to columnaris disease than the parental species. Diseases of aquatic organisms 100:77-81.

Arimura A, Shioda S (1995) Pituitary adenylate cyclase activating polypeptide (PACAP) and its receptors: neuroendocrine and endocrine interaction. Frontiers in neuroendocrinology 16:53-88.

Aubin I, Adams CP, Opsahl S, Septier D, Bishop CE, Auge N, Salvayre R, Negre-Salvayre A, Goldberg M, Guénet J-L (2005) A deletion in the gene encoding sphingomyelin phosphodiesterase 3 (Smpd3) results in osteogenesis and dentinogenesis imperfecta in the mouse. Nature genetics 37:803-805.

Beck BH, Farmer BD, Straus DL, Li C, Peatman E (2012) Putative roles for a rhamnose binding lectin in *Flavobacterium columnare* pathogenesis in channel catfish *Ictalurus punctatus*. Fish & shellfish immunology 33:1008-1015.

Birchler JA, Auger DL, Riddle NC (2003) In search of the molecular basis of heterosis. The Plant Cell Online 15:2236-2239.

Bodart V, Babinski K, Ong H, De Lean A (1997) Comparative Effect of Pituitary Adenylate Cyclase-Activating Polypeptide on Aldosterone Secretion in Normal Bovine and Human Tumorous Adrenal Cells 1. Endocrinology 138:566-573.

Bone H, Williams NA (2001) Antigen-receptor cross-linking and lipopolysaccharide trigger distinct phosphoinositide 3-kinase-dependent pathways to NF-kappa B activation in primary B cells. International Immunology 13:807-816.

Brock C, Schaefer M, Reusch HP, Czupalla C, Michalke M, Spicher K, Schultz G, Nurnberg B (2003) Roles of G beta gamma in membrane recruitment and activation of p110 gamma/p101 phosphoinositide 3-kinase gamma. The Journal of cell biology 160:89-99.

Burge C, Karlin S (1997) Prediction of complete gene structures in human genomic DNA. Journal of molecular biology 268:78-94.

Bush WS, Moore JH (2012) Genome-wide association studies. PLoS computational biology 8:e1002822.
Chockalingam PS, Gee SH, Jarrett HW (1999) Pleckstrin homology domain 1 of mouse α1-syntrophin binds phosphatidylinositol 4, 5-bisphosphate. Biochemistry 38:5596-5602.
Cox D, Tseng CC, Bjekic G, Greenberg S (1999) A requirement for phosphatidylinositol 3-kinase in pseudopod extension. Journal of Biological Chemistry 274:1240-1247.
Cummings HE, Barbi J, Reville P, Oghumu S, Zorko N, Sarkar A, Keiser TL, Lu B, Rückle T, Varikuti S (2012) Critical role for phosphoinositide 3-kinase gamma in parasite invasion and disease progression of cutaneous leishmaniasis. Proceedings of the National Academy of Sciences 109:1251-1256.
Darvasi A, Soller M (1992) Selective genotyping for determination of linkage between a marker locus and a quantitative trait locus. Theoretical and applied genetics 85:353-359.
Darvasi A, Soller M (1995) Advanced intercross lines, an experimental population for fine genetic mapping. Genetics 141:1199-1207.
Declercq AM, Haesebrouck F, Van den Broeck W, Bossier P, Decostere A (2013) Columnaris disease in fish: a review with emphasis on bacterium-host interactions. Veterinary research 44:27-44.
Decostere A, Haesebrouck F, Van Driessche E, Charlier G, Ducatelle R (1999) Characterization of the adhesion of Flavobacterium columnare (*Flexibacter columnaris*) to gill tissue. Journal of Fish Diseases 22:465-474.
Dominik S (2013) Descriptive Statistics of Data: Understanding the Data Set and Phenotypes of Interest. In Genome-Wide Association Studies and Genomic Prediction (Springer), pp. 19-35.
Downes C, Bennett D, McConnachie G, Leslie N, Pass I, MacPhee C, Patel L, Gray A (2001) Antagonism of PI 3-kinase-dependent signalling pathways by the tumour suppressor protein, PTEN. Biochemical Society Transactions 29:846-851.
Duggal P, Gillanders EM, Holmes TN, Bailey-Wilson JE (2008) Establishing an adjusted p-value threshold to control the family-wide type 1 error in genome wide association studies. BMC genomics 9:516.
Dunham RA, Umali GM, Beam R, Kristanto AH, Trask M (2008) 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 70:98-106.
Fukao T, Tanabe M, Terauchi Y, Ota T, Matsuda S, Asano T, Kadowaki T, Takeuchi T, Koyasu S (2002) PI3K-mediated negative feedback regulation of IL-12 production in DCs. Nature immunology 3:875-881.
Fulker D, Cherny S, Sham P, Hewitt J (1999) Combined linkage and association sib-pair analysis for quantitative traits. The American Journal of Human Genetics 64:259-267.
Gao J, Huo L, Sun X, Liu M, Li D, Dong J-T, Zhou J (2008) The tumor suppressor CYLD regulates microtubule dynamics and plays a role in cell migration. Journal of Biological Chemistry 283:8802-8809.
Gauderman W, Morrison J (2006) QUANTO 1.1: A computer program for power and sample size calculations for genetic-epidemiology studies.
Geng X, Sha J, Liu S, Bao L, Zhang J, Wang R, Yao J, Li C, Feng J, Sun F (2015) A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. BMC genomics 16:196.

Gheyas A, Haley C, Guy D, Hamilton A, Tinch A, Mota‐Velasco J, Woolliams J (2010a) Effect of a major QTL affecting IPN resistance on production traits in Atlantic salmon. Animal genetics 41:666-668.
Gheyas A, Houston R, Mota‐Velasco J, Guy D, Tinch A, Haley C, Woolliams J (2010b) Segregation of infectious pancreatic necrosis resistance QTL in the early life cycle of Atlantic Salmon (*Salmo salar*). Animal genetics 41:531-536.
Giudice JJ (1966) Growth of a blue X channel catfish hybrid as compared to its parent species. The Progressive Fish-Culturist 28:142-145.
Gu X, Feng C, Ma L, Song C, Wang Y, Da Y, Li H, Chen K, Ye S, Ge C (2011) Genome-wide association study of body weight in chicken F2 resource population. PLoS One 6:e21872.
Gudbjartsson DF, Walters GB, Thorleifsson G, Stefansson H, Halldorsson BV, Zusmanovich P, Sulem P, Thorlacius S, Gylfason A, Steinberg S (2008) Many sequence variants affecting diversity of adult human height. Nature genetics 40:609-615.
Hawke JP, Thune RL (1992) Systemic isolation and antimicrobial susceptibility of Cytophaga columnaris from commercially reared channel catfish. Journal of Aquatic Animal Health 4:109-113.
Hayes B (2013) Overview of statistical methods for genome-wide association studies (GWAS). In Genome-Wide Association Studies and Genomic Prediction (Springer), pp. 149-169.
Hirschhorn JN, Daly MJ (2005) Genome-wide association studies for common diseases and complex traits. Nature Reviews Genetics 6:95-108.
Hochberg Y, Benjamini Y (1990) More powerful procedures for multiple significance testing. Statistics in medicine 9:811-818.
Holleran EA, Ligon LA, Tokito M, Stankewich MC, Morrow JS, Holzbaur EL (2001) beta III spectrin binds to the Arp1 subunit of dynactin. The Journal of biological chemistry 276:36598-36605.
Houston R, Gheyas A, Hamilton A, Guy D, Tinch A, Taggart J, McAndrew B, Haley C, Bishop S (2008) Detection and confirmation of a major QTL affecting resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (Salmo salar). Developments in biologicals 132:199-204.
Houston RD, Davey JW, Bishop SC, Lowe NR, Mota-Velasco JC, Hamilton A, Guy DR, Tinch AE, Thomson ML, Blaxter ML (2012) Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. BMC genomics 13:244.
Howell GR, Shindo M, Murray S, Gridley T, Wilson LA, Schimenti JC (2007) Mutation of a ubiquitously expressed mouse transmembrane protein (Tapt1) causes specific skeletal homeotic transformations. Genetics 175:699-707.
Ireton K, Payrastre B, Chap H, Ogawa W, Sakaue H, Kasuga M, Cossart P (1996) A role for phosphoinositide 3-kinase in bacterial invasion. Science 274:780-782.
Ireton K, Payrastre B, Cossart P (1999) The Listeria monocytogenes protein InlB is an agonist of mammalian phosphoinositide 3-kinase. The Journal of biological chemistry 274:17025-17032.
Itzstein C, Coxon FP, Rogers MJ (2011) The regulation of osteoclast function and bone resorption by small GTPases. Small GTPases 2:117-130.
Jia G, Huang X, Zhi H, Zhao Y, Zhao Q, Li W, Chai Y, Yang L, Liu K, Lu H (2013) A haplotype map of genomic variations and genome-wide association studies of agronomic traits in foxtail millet (Setaria italica). Nature genetics 45:957-961.

Jiang K, Zhong B, Gilvary DL, Corliss BC, Hong-Geller E, Wei S, Djeu JY (2000) Pivotal role of phosphoinositide-3 kinase in regulation of cytotoxicity in natural killer cells. Nature immunology 1:419-425.
Kang HM, Sul JH, Service SK, Zaitlen NA, Kong S-y, Freimer NB, Sabatti C, Eskin E (2010) Variance component model to account for sample structure in genome-wide association studies. Nature genetics 42:348-354.
Kang HM, Zaitlen NA, Wade CM, Kirby A, Heckerman D, Daly MJ, Eskin E (2008) Efficient control of population structure in model organism association mapping. Genetics 178:1709-1723.
Kierbel A, Gassama-Diagne A, Mostov K, Engel JN (2005) The phosphoinositol-3-kinase-protein kinase B/Akt pathway is critical for *Pseudomonas aeruginosa* strain PAK internalization. Molecular biology of the cell 16:2577-2585.
Korte A, Farlow A (2013) The advantages and limitations of trait analysis with GWAS: a review. Plant methods 9:29.
Koyasu S (2006) Role of phosphatidylinositol 3-kinase in the immune system. Tanpakushitsu kakusan koso Protein, nucleic acid, enzyme 51:1569-1579.
Kucuktas H, Wang S, Li P, He C, Xu P, Sha Z, Liu H, Jiang Y, Baoprasertkul P, Somridhivej B (2009) Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics 181:1649-1660.
LaFrentz BR, Shoemaker CA, Booth NJ, Peterson BC, Ourth DD (2012) 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 24:141-147.
Laghari M, Lashari P, Zhang X, Xu P, Xin B, Zhang Y, Narejo N, Sun X (2014) Mapping quantitative trait loci (QTL) for body weight, length and condition factor traits in backcross (BC1) family of Common carp (Cyprinus carpio L.). Molecular biology reports 41:721-731.
Laird NM, Lange C (2006) Family-based designs in the age of large-scale gene-association studies. Nature Reviews Genetics 7:385-394.
Lambotin M, Hoffmann I, Laran-Chich MP, Nassif X, Couraud PO, Bourdoulous S (2005) Invasion of endothelial cells by *Neisseria meningitidis* requires cortactin recruitment by a phosphoinositide-3-kinase/Rac1 signalling pathway triggered by the lipo-oligosaccharide. Journal of cell science 118:3805-3816.
Lange C, DeMeo DL, Laird NM (2002) Power and design considerations for a general class of family-based association tests: quantitative traits. The American Journal of Human Genetics 71:1330-1341.
Lasky-Su J, Won S, Mick E, Anney RJ, Franke B, Neale B, Biederman J, Smalley SL, Loo SK, Todorov A (2010) On genome-wide association studies for family-based designs: an integrative analysis approach combining ascertained family samples with unselected controls. The American Journal of Human Genetics 86:573-580.
Ledur M, Navarro N, Pérez-Enciso M (2009) Large-scale SNP genotyping in crosses between outbred lines: how useful is it&quest. Heredity 105:173-182.
Lewontin R (1964) The interaction of selection and linkage. I. General considerations; heterotic models. Genetics 49:49.
Li C, Zhang Y, Wang R, Lu J, Nandi S, Mohanty S, Terhune J, Liu Z, Peatman E (2012) 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 32:816-827.
Li Y, Liu S, Qin Z, Waldbieser G, Wang R, Sun L, Bao L, Danzmann RG, Dunham R, Liu Z (2014) Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. DNA research : an international journal for rapid publication of reports on genes and genomes:dsu038.
Liu F, van der Lijn F, Schurmann C, Zhu G, Chakravarty MM, Hysi PG, Wollstein A, Lao O, de Bruijne M, Ikram MA (2012) A genome-wide association study identifies five loci influencing facial morphology in Europeans.
Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y, Zhang J, Feng J, Kaltenboeck L, Kucuktas H (2014) Development of the catfish 250K SNP array for genome-wide association studies. BMC research notes 7:135.
Liu S, Zhou Z, Lu J, Sun F, Wang S, Liu H, Jiang Y, Kucuktas H, Kaltenboeck L, Peatman E, Liu Z (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. BMC genomics 12:53-66.
Liu Z, Karsi A, Li P, Cao D, Dunham R (2003) An AFLP-based genetic linkage map of channel catfish (Ictalurus punctatus) constructed by using an interspecific hybrid resource family. Genetics 165:687-694.
Longtine MS, Bi E (2003) Regulation of septin organization and function in yeast. Trends in cell biology 13:403-409.
Lugo JM, Carpio Y, Oliva A, Morales A, Estrada MP (2010) Pituitary adenylate cyclase-activating polypeptide (PACAP): a regulator of the innate and acquired immune functions in juvenile fish. Fish & shellfish immunology 29:513-520.
Luo C, Qu H, Ma J, Wang J, Li C, Yang C, Hu X, Li N, Shu D (2013) Genome-wide association study of antibody response to Newcastle disease virus in chicken. BMC Genetics 14:42-51.
Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits, Vol 1 (Sinauer Sunderland).
Mackay I, Powell W (2007) Methods for linkage disequilibrium mapping in crops. Trends in plant science 12:57-63.
Mackay TF (2001a) The genetic architecture of quantitative traits. Annual review of genetics 35:303-339.
Mackay TF (2001b) Quantitative trait loci in *Drosophila*. Nature Reviews Genetics 2:11-20.
Maekawa T, Krauss JL, Abe T, Jotwani R, Triantafilou M, Triantafilou K, Hashim A, Hoch S, Curtis MA, Nussbaum G (2014) *Porphyromonas gingivalis* Manipulates Complement and TLR Signaling to Uncouple Bacterial Clearance from Inflammation and Promote Dysbiosis. Cell host & microbe 15:768-778.
Maga AM, Navarro N, Cunningham ML, Cox TC (2015) Quantitative trait loci affecting the 3D skull shape and size in mouse and prioritization of candidate genes in-silico. Frontiers in physiology 6.
Massault C, Hellemans B, Louro B, Batargias C, Van Houdt J, Canario A, Volckaert F, Bovenhuis H, Haley C, De Koning D (2010) QTL for body weight, morphometric traits and stress response in European sea bass Dicentrarchus labrax. Animal Genetics 41:337-345.
Massoumi R (2010) Ubiquitin chain cleavage: CYLD at work. Trends Biochem Sci 35:392-399.
McPherron AC, Lawler AM, Lee S-J (1999) Regulation of anterior/posterior patterning of the axial skeleton by growth/differentiation factor 11. Nature genetics 22:260-264.

Michalak P (2008) Coexpression, coregulation, and cofunctionality of neighboring genes in eukaryotic genomes. Genomics 91:243-248.
Michelmore RW, Paran I, Kesseli R (1991) 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 88:9828-9832.
Moen T, Baranski M, Sonesson AK, Kjøglum S (2009) Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*): population-level associations between markers and trait. BMC genomics 10:368.
Morin PA, Luikart G, Wayne RK (2004) SNPs in ecology, evolution and conservation. Trends in Ecology & Evolution 19:208-216.
Mott R, Talbot CJ, Turri MG, Collins AC, Flint J (2000) A method for fine mapping quantitative trait loci in outbred animal stocks. Proceedings of the National Academy of Sciences 97:12649-12654.
Muffato M, Louis A, Poisnel C-E, Crollius HR (2010) Genomicus: a database and a browser to study gene synteny in modern and ancestral genomes. Bioinformatics 26:1119-1121.
Ninwichian P, Peatman E, Liu H, Kucuktas H, Somridhivej B, Liu S, Li P, Jiang Y, Sha Z, Kaltenboeck L (2012) Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map. G3: Genes| Genomes| Genetics 2:1233-1241.
Nishimura S, Watanabe T, Mizoshita K, Tatsuda K, Fujita T, Watanabe N, Sugimoto Y, Takasuga A (2012) Genome-wide association study identified three major QTL for carcass weight including the PLAG1-CHCHD7 QTN for stature in Japanese Black cattle. BMC genetics 13:40.
Olivares-Fuster O, Arias CR (2011) Development and characterization of rifampicin-resistant mutants from high virulent strains of Flavobacterium columnare. Journal of fish diseases 34:385-394.
Olivares-Fuster O, Bullard SA, McElwain A, Llosa MJ, Arias CR (2011) Adhesion dynamics of Flavobacterium columnare to channel catfish *Ictalurus punctatus* and zebrafish *Danio rerio* after immersion challenge. Diseases of aquatic organisms 96:221-227.
Palti Y, Gao G, Liu S, Kent M, Lien S, Miller M, Rexroad C, Moen T (2014) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. Molecular ecology resources.
Peatman E, Li C, Peterson BC, Straus DL, Farmer BD, Beck BH (2013) Basal polarization of the mucosal compartment in *Flavobacterium columnare* susceptible and resistant channel catfish (*Ictalurus punctatus*). Molecular immunology 56:317-327.
Phillips R, Ventura A, Dekoning J, Nichols K (2013) Mapping rainbow trout immune genes involved in inflammation reveals conserved blocks of immune genes in teleosts. Animal genetics 44:107-113.
Pizarro-Cerda J, Cossart P (2006) Bacterial adhesion and entry into host cells. Cell 124:715-727.
Plumb JA, Hanson LA, Plumb JA (2011) Health maintenance and principal : microbial diseases of cultured fishes, 3rd edn (Ames, Iowa, Wiley-Blackwell).
Price AL, Patterson NJ, Plenge RM, Weinblatt ME, Shadick NA, Reich D (2006) Principal components analysis corrects for stratification in genome-wide association studies. Nature genetics 38:904-909.
Price AL, Zaitlen NA, Reich D, Patterson N (2010) New approaches to population stratification in genome-wide association studies. Nature Reviews Genetics 11:459-463.

Pritchard JK, Przeworski M (2001) Linkage disequilibrium in humans: models and data. The American Journal of Human Genetics 69:1-14.
Pruitt KD, Tatusova T, Maglott DR (2007) NCBI reference sequences (RefSeq): a curated non-redundant sequence database of genomes, transcripts and proteins. Nucleic acids research 35:D61-D65.
Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MA, Bender D, Maller J, Sklar P, De Bakker PI, Daly MJ (2007) PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81:559-575.
Quarrie SA, Lazić-Jančić V, Kovačević D, Steed A, Pekić S (1999) Bulk segregant analysis with molecular markers and its use for improving drought resistance in maize. Journal of experimental botany 50:1299-1306.
Quilliam LA, Rebhun JF, Castro AF (2002) A growing family of guanine nucleotide exchange factors is responsible for activation of Ras-family GTPases. Progress in nucleic acid research and molecular biology 71:391-444.
Reddy SAG, Huang JH, Liao WSL (1997) Phosphatidylinositol 3-kinase in interleukin 1 signaling - Physical interaction with the interleukin 1 receptor and requirement in NF kappa B and AP-1 activation. Journal of Biological Chemistry 272:29167-29173.
Reich DE, Cargill M, Bolk S, Ireland J, Sabeti PC, Richter DJ, Lavery T, Kouyoumjian R, Farhadian SF, Ward R (2001) Linkage disequilibrium in the human genome. Nature 411:199-204.
Risch N, Merikangas K (1996) The future of genetic studies of complex human diseases. Science 273:1516-1517.
Romanelli F, Fillo S, Isidori A, Conte D (1997) Pituitary adenylate cyclase-activating polypeptide regulates rat Leydig cell function in vitro. Neuropeptides 31:311-317.
Salamov AA, Solovyev VV (2000) Ab initio gene finding in Drosophila genomic DNA. Genome research 10:516-522.
Schmid M, Davison TS, Henz SR, Pape UJ, Demar M, Vingron M, Schölkopf B, Weigel D, Lohmann JU (2005) A gene expression map of *Arabidopsis thaliana* development. Nature genetics 37:501-506.
Schoenebeck JJ, Hutchinson SA, Byers A, Beale HC, Carrington B, Faden DL, Rimbault M, Decker B, Kidd JM, Sood R (2012) Variation of BMP3 contributes to dog breed skull diversity. PLoS Genet 8:e1002849.
Schoenebeck JJ, Ostrander EA (2013) The genetics of canine skull shape variation. Genetics 193:317-325.
Sémon M, Duret L (2006) Evolutionary origin and maintenance of coexpressed gene clusters in mammals. Molecular biology and evolution 23:1715-1723.
Shears SB (2009) Diphosphoinositol polyphosphates: metabolic messengers? Molecular pharmacology 76:236-252.
Sidis Y, Tortoriello DV, Holmes WE, Pan Y, Keutmann HT, Schneyer AL (2002) Follistatin-related protein and follistatin differentially neutralize endogenous vs. exogenous activin. Endocrinology 143:1613-1624.
Sjöström M, Johansson A-S, Schröder O, Qiu H, Palmblad J, Haeggström JZ (2003) Dominant expression of the CysLT2 receptor accounts for calcium signaling by cysteinyl leukotrienes in human umbilical vein endothelial cells. Arteriosclerosis, thrombosis, and vascular biology 23:e37-e41.

Smitherman RO, Dunham RA, Tave D (1983) Review of catfish breeding research 1969–1981 at Auburn University. Aquaculture 33:197-205.
Solovyev V, Kosarev P, Seledsov I, Vorobyev D (2006) Automatic annotation of eukaryotic genes, pseudogenes and promoters. Genome Biol 7:S10.
Spielman RS, McGinnis RE, Ewens WJ (1993) Transmission test for linkage disequilibrium: the insulin gene region and insulin-dependent diabetes mellitus (IDDM). American journal of human genetics 52:506.
Sun F, Peatman E, Li C, Liu S, Jiang Y, Zhou Z, Liu Z (2012) Transcriptomic signatures of attachment, NF-kappaB suppression and IFN stimulation in the catfish gill following columnaris bacterial infection. Developmental and comparative immunology 38:169-180.
Sun L, Liu S, Wang R, Jiang Y, Zhang Y, Zhang J, Bao L, Kaltenboeck L, Dunham R, Waldbieser G (2014) Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (Ictalurus punctatus). PloS one 9:e109666.
Sun SC (2010) CYLD: a tumor suppressor deubiquitinase regulating NF-kappa B activation and diverse biological processes. Cell Death and Differentiation 17:25-34.
Sutter NB, Bustamante CD, Chase K, Gray MM, Zhao K, Zhu L, Padhukasahasram B, Karlins E, Davis S, Jones PG (2007) A single IGF1 allele is a major determinant of small size in dogs. Science 316:112-115.
Szebenyi G, Fallon JF (1998) Fibroblast growth factors as multifunctional signaling factors. International review of cytology 185:45-106.
Terrien E, Chaffotte A, Lafage M, Khan Z, Prehaud C, Cordier F, Simenel C, Delepierre M, Buc H, Lafon M (2012) Interference with the PTEN-MAST2 interaction by a viral protein leads to cellular relocalization of PTEN. Science signaling 5:ra58.
Thompson C, Cloutier A, Bossé Y, Poisson C, Larivée P, McDonald PP, Stankova J, Rola-Pleszczynski M (2008) Signaling by the Cysteinyl-Leukotriene Receptor 2 involvement in chemokine gene transcription. Journal of Biological Chemistry 283:1974-1984.
Tomiuk S, Hofmann K, Nix M, Zumbansen M, Stoffel W (1998) Cloned mammalian neutral sphingomyelinase: Functions in sphingolipid signaling? Proceedings of the National Academy of Sciences 95:3638-3643.
Turner S, Armstrong LL, Bradford Y, Carlson CS, Crawford DC, Crenshaw AT, Andrade M, Doheny KF, Haines JL, Hayes G (2011) Quality control procedures for genome‐wide association studies. Current protocols in human genetics:1.19. 11-11.19. 18.
Van Gestel S, Houwing-Duistermaat JJ, Adolfsson R, van Duijn CM, Van Broeckhoven C (2000) Power of selective genotyping in genetic association analyses of quantitative traits. Behavior genetics 30:141-146.
Viard P, Butcher AJ, Halet G, Davies A, Nürnberg B, Heblich F, Dolphin AC (2004) PI3K promotes voltage-dependent calcium channel trafficking to the plasma membrane. Nature neuroscience 7:939-946.
Vieira OV, Botelho RJ, Rameh L, Brachmann SM, Matsuo T, Davidson HW, Schreiber A, Backer JM, Cantley LC, Grinstein S (2001) Distinct roles of class I and class III phosphatidylinositol 3-kinases in phagosome formation and maturation. The Journal of cell biology 155:19-25.
Viel A, Branton D (1996) Spectrin: on the path from structure to function. Current opinion in cell biology 8:49-55.

Wang J, Yang G, Zhou G (2013a) Quantitative trait loci for morphometric body measurements of the hybrids of silver carp (Hypophthalmichthys molitrix) and bighead carp (H. nobilis). Acta Biologica Hungarica 64:169-183.
Wang KZ, Wara-Aswapati N, Boch JA, Yoshida Y, Hu C-D, Galson DL, Auron PE (2006) TRAF6 activation of PI 3-kinase-dependent cytoskeletal changes is cooperative with Ras and is mediated by an interaction with cytoplasmic Src. Journal of cell science 119:1579-1591.
Wang R, Sun L, Bao L, Zhang J, Jiang Y, Yao J, Song L, Feng J, Liu S, Liu Z (2013b) Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. BMC genomics 14:929.
Wang WY, Barratt BJ, Clayton DG, Todd JA (2005) Genome-wide association studies: theoretical and practical concerns. Nature Reviews Genetics 6:109-118.
Watanabe T, Sato K, Kaibuchi K (2009) Cadherin-mediated intercellular adhesion and signaling cascades involving small GTPases. Cold Spring Harbor perspectives in biology 1:a003020.
Williams EJ, Bowles DJ (2004) Coexpression of neighboring genes in the genome of Arabidopsis thaliana. Genome Research 14:1060-1067.
Won S, Wilk JB, Mathias RA, O'Donnell CJ, Silverman EK, Barnes K, O'Connor GT, Weiss ST, Lange C (2009) On the analysis of genome-wide association studies in family-based designs: a universal, robust analysis approach and an application to four genome-wide association studies. PLoS genetics 5:e1000741.
Wu H, Arron JR (2003) TRAF6, a molecular bridge spanning adaptive immunity, innate immunity and osteoimmunology. Bioessays 25:1096-1105.
Wunnenberg-Stapleton K, Blitz IL, Hashimoto C, Cho K (1999) Involvement of the small GTPases XRhoA and XRnd1 in cell adhesion and head formation in early Xenopus development. Development 126:5339-5351.
Xiong H, Li H, Chen Y, Zhao J, Unkeless JC (2004) Interaction of TRAF6 with MAST205 regulates NF-kappaB activation and MAST205 stability. The Journal of biological chemistry 279:43675-43683.
Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y, Kuang Y, Zhang Y, Feng J, Li C (2014) Development and evaluation of the first high-throughput SNP array for common carp (Cyprinus carpio). BMC genomics 15:307.
Yang J, Lee SH, Goddard ME, Visscher PM (2011) GCTA: a tool for genome-wide complex trait analysis. The American Journal of Human Genetics 88:76-82.
Yang J, Zaitlen NA, Goddard ME, Visscher PM, Price AL (2014) Advantages and pitfalls in the application of mixed-model association methods. Nature genetics 46:100-106.
Yang W-L, Jin G, Li C-F, Jeong YS, Moten A, Xu D, Feng Z, Chen W, Cai Z, Darnay B (2013) Cycles of ubiquitination and deubiquitination critically regulate growth factor-mediated activation of Akt signaling. Science signaling 6:ra3.
Yu J, Pressoir G, Briggs WH, Bi IV, Yamasaki M, Doebley JF, McMullen MD, Gaut BS, Nielsen DM, Holland JB (2005) A unified mixed-model method for association mapping that accounts for multiple levels of relatedness. Nature genetics 38:203-208.
Zhang Z, Ersoz E, Lai C-Q, Todhunter RJ, Tiwari HK, Gore MA, Bradbury PJ, Yu J, Arnett DK, Ordovas JM (2010) Mixed linear model approach adapted for genome-wide association studies. Nature genetics 42:355-360.
Zhou X, Stephens M (2012) Genome-wide efficient mixed-model analysis for association studies. Nature genetics 44:821-824.

Ziegler A, König IR, Thompson JR (2008) Biostatistical Aspects of Genome - Wide Association Studies. Biometrical Journal 50:8-28.