

Development of 690K SNP arrays for whole genome mapping and genetic studies in catfish

by

Qifan Zeng

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

Auburn, Alabama
August 6, 2016

Keywords: catfish, genome, SNP array, linkage map, next generation sequencing

Copyright 2016 by Qifan Zeng

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
Nannan Liu, Professor of Insect Molecular Toxicology and Physiology
Charles Y. Chen, Associate Professor of Crop, Soil and Environmental Sciences

Abstract

Catfish is the leading aquaculture species in the United States, accounting for approximately 60% of U.S. aquaculture production. However, the catfish industry is facing great challenges including escalating feed and fuel costs, intense international competition, and devastating diseases. Application of genetic markers facilitate elucidating the genetic basis for quantitative traits, and enabling characterization of the major quantitative trait loci (QTLs) and underlying causative genes. Efficient marker systems need to be developed to allow automated and efficient genotyping at an affordable cost. In catfish, the first SNP array with previously identified 250,113 SNPs have been developed to facilitate the analysis of quantitative traits. It has been developed by using Axiom genotyping technology which provides a flexible solution to design arrays with customizable markers from non-model organisms. Recently, the advancement of array technologies improved the performance of the Axiom genotyping arrays with expended capacities, enabling the construction of high density SNP arrays.

In the first chapter of this dissertation, I will provide detailed description of a catfish 690K SNP arrays with the Axiom genotyping platform. The catfish 690K SNP array includes 690,662 unique SNPs markers which were identified from whole genome sequencing, RNA-seq, and genotyping by sequencing (GBS) data over 1,200 catfish. With the guidance of channel catfish reference genome sequence, the selected SNPs were relatively evenly distributed across the entire genome, with 0.8 kb interquartile range of intervals, and covered 99% of the genome scaffolds and 93.9% of the BAC-based physical map contigs. With enhanced marker density and power of

discrimination, the catfish 690K SNP array should be a valuable tool for catfish research community and greatly improve the genetic and genomic studies.

In the second chapter of this dissertation, I will describe the application of the 690K SNP array for the construction of a very high density genetic linkage map. Construction of accurate and high density genetic linkage map is fundamental for genetic and genomic studies. In the present work, I constructed a new generation of channel catfish genetic linkage map with improved marker density and resolution with four large reference families genotyped using the catfish 690K SNP array. Over 250,000 SNPs were finally placed on the linkage map, which contains, to the best of our knowledge, the highest marker density among all the aquaculture species. The total estimated genetic size of the new linkage map is 3,004 cM with a resolution of 0.1 cM for sex-averaged genetic map. The sex-specific linkage maps constructed in each sex revealed that the recombination rate is 1.4-fold larger in females than in males. After integration with the recently constructed catfish reference genome assembly, over 1,500 of the whole-genome sequencing contigs were anchored to the linkage groups, covering a physical length of 766 Mb that account for about 97.8% of the total genome assembly. Over 900 previously unmapped scaffolds were placed onto the linkage map, which greatly improves the catfish reference genome assembly. With the unprecedented coverage, it should also serve as a valuable tool for genetic analysis, especially GWAS and fine-scale QTL mapping for genes associated with economically important traits.

Acknowledgments

I would like to express my sincere gratitude and deep appreciation to my major professor, Dr. Zhanjiang Liu, for his support and guidance in my Ph.D. study. I am grateful for the time and expertise offered by my committee members, Dr. Rex Dunham, Dr. Nannan Liu, and Dr. Charles Y. Chen, and to thank my outside reader Dr. Luxin Wang. Furthermore, I would like to extend my sincere appreciation to Dr. Shikai Liu, Dr. Ruijia Wang, Dr. Chao Li, Dr. Luyang Sun, Dr. Fanyue Sun, Dr. Jiaren Zhang, Dr. Yu Zhang, Dr. Yun Li, Dr. Jun Yao, Dr. Chen Jiang, Dr. Xin Geng, Dr. Lisui Bao, Dr. Ailu Chen and Dr. Lin Song for their guidance and encouragement. My thanks also go to Qiang Fu, Sen Gao, Tao Zhou, Ning Li, Xiaozhu Wang, Zihao Yuan, Yujia Yang, Yulin Jin, Suxu Tan, Changxu Tian, Xiaoxiao Zhong and all the other colleagues in the laboratory for their help and friendship. I am grateful to the Chinese Scholarship Council for the financial support.

Above all, I want to thank my father Yongqing Zeng, my mother Hui Wang, and my fiancée Xiaohan Zhang for their constant love and support.

Table of Contents

Abstract.....	ii
Acknowledgments	ii
List of Tables	v
List of Figures.....	vi
CHAPTER 1 INTRODUCTION	1
Overview	1
Genetic breeding in catfish.....	3
Genome wide SNP discovery.....	5
Development of SNP genotyping array in agriculture and aquaculture species	7
CHAPTER 2 GENOME-WIDE SNP IDENTIFICATION AND DEVELOPMENT OF A HIGH-DENSITY SNP ARRAY	12
Abstract	12
Introduction	13
Materials and methods	15
Results and discussion.....	21
Conclusion.....	32
CHAPTER 3 CONSTRUCTION OF A HIGH-DENSITY GENETIC MAP AND ITS INTEGRATION WITH GENOMIC SEQUENCES OF CHANNEL CATFISH.....	34
Abstract	34
Introduction	35

Materials and methods	38
Results	42
Discussion	54
REFERENCES	57

List of Tables

Table 1. Sample and data size for development of 690K SNP array.....	15
Table 2. Summary of SNPs identified in channel and blue catfish.	23
Table 3. Summary of SNPs identified in channel catfish domestic strains and wild population.	23
Table 4. Summary of the catfish 690K SNP array.	25
Table 5. Summary of the species/strain-specific SNPs on catfish 690K SNP array.	28
Table 6. SNP metrics summary.	32
Table 7. Summary of the sex-average linkage map of channel catfish.	45
Table 8. Summary of the sex-specific linkage map of channel catfish.	48

List of Figures

Figure 1. Distribution of SNP probes based on p-conver values.....	26
Figure 2. Distribution of SNPs based on MAF.....	27
Figure 3. Genome distribution of SNPs on the 690K SNP array.	29
Figure 4. Distribution of inter-SNPs spacing for SNPs included on the array.	31
Figure 5. Box-plot of inter-SNPs spacing for SNPs included on the array.	31
Figure 6. Genetic mapping procedure.....	42
Fig 7. Sample structure identified by multidimensional scaling analysis of IBS distances.	43
Figure 8. SNPs filtering process for linkage mapping.....	44
Figure 9. Illustration of sex-specific linkage map	47
Figure 10. Illustration of female-specific linkage map.....	50
Figure 10. Illustration of male-specific linkage map.....	51
Figure 11. Concordance of SNP marker positions on reference sequence with those on genetic linkage map.	53

CHAPTER 1. INTRODUCTION

Overview

Channel catfish (*Ictalurus punctatus*) is one of the most important aquaculture species in the United States of America. They were naturally distributed across eastern and northern United States, as well as northern Mexico. Domestication of channel catfish was started over 100 years ago (Leary 1908). Since the ictalurid genetic evaluation and early breeding programs in the 1950s and 1960s, channel catfish has been identified as the most suitable species of catfish for aquaculture (Dunham and Smitherman 1984). Commercial aquaculture of catfish in the United States was considered to start in the late 1950s (Swingle 1958) and developed rapidly since 1960s as the consequence of improvements in disease control, feed supply, and pond management (Smitherman and Dunham 1993). To date, channel catfish farming has been introduced to over 13 countries worldwide, with approximately 800 million pounds of global annual production (FAO Global Aquaculture Production Statistics 1950-2013). As a result of domestication process for past decades, many domesticated strains have been established with distinct adaptations to farmed environments and economic traits to meet consumer demands. The United States is one of the largest channel catfish producers, with various domesticated strains and breeds, including Kansas, Marion, Thompson, USDA103, Hatchery, Arkansas, as well as many channel-blue catfish hybrid strains, all of which were the basis and genetic resources for selective breeding using modern genetic tools.

However, catfish industry, the leading aquaculture industry in the U.S. have encountered great challenges. Due to escalating catfish feed and fuel costs, intense international competition, and devastating diseases, the annual production of channel catfish in the U.S. decreased significantly since a high mark in 2003 when over 600 million pounds of round weight catfish were processed. In 2014, only 301 million pounds were processed, decreased 54% since the 2003 peak. The shortage in catfish supply was made up by imported exotic catfish fillets which account for 80% of all U.S. sales of frozen catfish and catfish-like fillet products (Hanson 2015). Until the production efficiency of domestic catfish is improved, it is likely that the industry will be further threatened by imported exotic catfishes.

Exploiting whole genome marker-based genomic selection can greatly contribute to improving performance traits and production efficiency, such as disease resistance, feed conversion rate, as well as hypoxic adaptation. Single nucleotide polymorphisms (SNPs) were now the choice of genetic markers as a consequence of the advancement of sequencing power and improved computational capacities in analyzing single nucleotide variations on a genome-wide scale (Helyar et al. 2011). In channel catfish, many efforts and progresses have been made to identify SNPs on genome-wide (He et al. 2003; Liu et al. 2011; Sun et al. 2014). Additionally, accumulated catfish sequencing data from previous experiments also provide a valuable resource for SNP markers development. With the comprehensive SNP data and guidance of reference genome sequence of channel catfish, a new generation of SNP array could be built for genetic analysis and whole genome marker-based selection.

The first objective of my study was to construct in a high-density and high-coverage SNP array by using a comprehensive catfish NGS datasets. In addition to this primary goal, the second goal was to construct a high-density genetic linkage map by using the newly designed SNP array, and to increase integration of the linkage and physical maps with the markers derived from BAC end sequences for genetic linkage analysis. This study will benefit genetic improvement of catfish breeding programs and serve as a valuable framework for validating the reference whole-genome sequences, extensive comparative and functional genomic studies, and fine-scale QTL mapping and association studies in catfish.

Genetic breeding in catfish

Since the catfish genetics and breeding program in the 1950s, channel catfish has been identified as the most suitable species for aquaculture compared with other catfish species, such as blue catfish (*Ictalurus furcatus*), white catfish (*Ameiurus catus*), brown bullhead catfish (*A. nebulosus*), and etc. (Dunham and Smitherman 1984). Later in 1960s, directed breeding was conducted via interspecific hybridization of seven major ictalurid species. Channel catfish female and blue catfish male produce heterotic hybrid catfish with remarkable performance (Dupree and Green 1969; Dupree et al. 1969; Giudice 1966). Efforts have also been made in channel catfish intraspecific crossbreeding and mass selection (Dunham and Smitherman 1982; Green et al. 1976), remarkable genetic improvement of catfish had been accomplished with 2 to 6% increased growth rate per generation (Dunham and Smitherman 1983). Additionally, triploid (Lilyestrom et al. 1999) and monosex production (Goudie et al. 1995) were evaluated, however, they did not show

much improvements and were abandoned due to lack of feasibility. As a result of decades of research efforts, progress has been made in mass selection, family selection, intraspecific crossbreeding, and interspecific introgression. Many domesticated strains have been established with significant variations in economic traits, including growth rate, feed conversion rate, disease resistance, seinability, and adaptations to farmed environments (Argue et al. 2003; Dunham 2007; Dunham and Argue 1998;2000). However, the effects of traditional selective breeding were limited by their relatively low intensity and accuracy. In contrast, molecular genetics hold promise for more widespread adoption with increased genetic gain and provide insight into the genomic levels of organization of inheritance traits (Eggen 2012; Goddard and Hayes 2009; Hayes and Goddard 2001).

To date, many efforts have also been made in catfish molecular genetics and genomics studies. A large number of genetic markers have been developed in channel and blue catfish, including microsatellites (Liu et al. 1999; Waldbieser and Bosworth 1997), AFLP (Liu et al. 1998), as well as SNPs (Liu et al. 2011; Sun et al. 2014). Several genetic linkage maps have been constructed with different types of molecular markers and multiple breeding families (Kucuktas et al. 2009; Li et al. 2015; Liu et al. 2003; Ninwichian et al. 2012; Waldbieser et al. 2001). A large set of expressed sequence tags (ESTs) have been generated using traditional cloning and Sanger sequencing methods, and next generation sequencing (Wang et al. 2010). A cDNA microarray has also been designed and constructed (Li and Waldbieser 2006). Recently, the channel catfish genome consortium has completely sequenced and assembled a reference genome sequence (data unpublished). All of these researches provide a valuable genetic resources for selective breeding assisted

by genome-wide markers, which will hopefully bring the upcoming genetic enhancement on the horizon and impact the catfish industry.

Genome wide SNP discovery

SNP represents a variation of a single base pair in the genomic DNA (Wang et al. 1998), which have been regarded as the marker of choice for many applications in genetics and genomics due to their suitability for automation, high abundance, and genome-wide coverage (Frazer et al. 2009; Morin et al. 2004). Theoretically, each SNP locus can possess four possible alleles since there were four types of nucleotide, but in most cases, the majority of SNPs have only two variants (Lai 2001). Analysis of biallelic SNP markers can be easily automated compared with multiallelic markers such as microsatellites and copy number variations (CNVs). The biallelic SNPs can result from either nucleotide transitions which were exchanges within purine or pyrimidine, or transversions which were exchanges between purine and pyrimidine. Owing to high spontaneous rate of deamination of 5-methylcytosine to thymidine, higher frequency of transition is observed than transversion (Holliday and Grigg 1993). SNPs were the most common type of genomic variations, about 90% of sequence variants were single base substitutions in humans (Collins et al. 1998). Depending on the positions of SNPs in the genomic DNA, single nucleotide substitutions can have different functional effects. SNPs in the coding regions were classified as exonic and intronic SNPs, which can result in changes in mRNA sequences and secondary structure formation (Cargill et al. 1999). SNPs in the noncoding regions can also affect transcription factor binding, gene splicing, and non-coding RNA sequences (Chen and Rajewsky 2006). The pronounced effects of SNPs on protein function and gene expression

provide a perspective for development of diagnostic SNP markers (Liu et al. 2006; Sattarzadeh et al. 2006).

Methods for SNP mining have evolved drastically. In early studies, large scale SNPs can be discovered via methods such as overlapping genomic DNA sequences (Taillon-Miller et al. 1998), aligning expressed sequence tags (ESTs) (Picoult-Newberg et al. 1999), and targeted re-sequencing of BAC-end sequences (Lorenz et al. 2010). These methods were either limited to model organisms with reference genome sequences, or were not reliable due to inadequate sequence depth (Boussaha et al. 2012; Wang et al. 2008). The advancement of next generation sequencing (NGS) technologies facilitated the construction of reference genome sequences and propelled the discovery of genome-wide SNP variations (Mardis 2008). Of all the NGS platforms, Illumina sequencing systems were demonstrated more effective for whole genome-wide SNP discovery because of the high-throughput (Eck et al. 2009) (Liu et al. 2012). With improvements on the patterned flow cell technology, the latest Illumina HiSeq system, HiSeq 4000, can generate over 750 Gb data per flow cell in a single run.

Multitude of studies have applied NGS platforms to acquire high-confidence SNPs on the whole-genome wide, especially in plants and animals with complex genomes (Stothard et al. 2011; You et al. 2011). For instance, cattle was one of the first agriculture species to enter the genomic era. In August 2006, the cattle reference genome sequence was released by the Human Genome Sequencing Center at Baylor College of Medicine, with more than 2.2 million putative SNPs (Elsik et al. 2009; Gibbs et al. 2009). Later, the Bovine Genome Sequencing Consortium screened out 62,042 putative SNPs from Illumina sequence reads of 66 cattle. Genotype data of selected SNPs presented a high validation

rate of 92% (Van Tassell et al. 2008). Gene-based SNPs in cattle were also discovered by using NGS data. In Djari et al.'s work, mRNA of three Limousin bull calves were subjected to Illumina high-throughput sequencing, identifying 8,407 high-confidence SNPs from 125 million reads (Djari et al. 2013). To date, large collections of SNPs have also been discovered in several other agricultural animals, including sheep (Kijas et al. 2009), chicken (Marklund and Carlborg 2010), swine (Ramos et al. 2009; Wiedmann et al. 2008), turkey (Aslam et al. 2012; Kerstens et al. 2009) and horse (Wade et al. 2009). Much efforts have been made for SNP mining in aquaculture species as well. In catfish, SNPs were identified from a wide range of studies. Beside SNPs identified from ESTs alignment (He et al. 2003; Wang et al. 2008), large numbers of SNPs were discovered from RNA-Seq (Liu et al. 2011), GBS (Li et al. 2014), and whole genome sequencing analysis (Sun et al. 2014). Extensive SNP datasets were also being generated in other aquaculture species, including Atlantic Herring (Helyar et al. 2012), common carp (Xu et al. 2012), plateau fish (Wang et al. 2015b), tiger puffer (Cui et al. 2014), rainbow trout (Palti et al. 2014), and blunt snout bream (Gao et al. 2012). The comprehensive SNPs database provide valuable resources for molecular marker assisted selection and whole genome selection (Eggen 2012).

Development of SNP genotyping array in agriculture and aquaculture species

The binary nature of the SNPs has made them highly suitable for automated genotyping methods. The most efficient approaches in high-throughput studies at this time include GBS and SNP array technologies. GBS can be processed with whole genome sequencing (Huang et al. 2009), target enrichment sequencing (Zhou and Holliday 2012),

and reduced representation sequencing strategies (Kumar et al. 2012). The decreasing cost of NGS were allowing it to be widely used to genotype large-scale SNPs for genetic mapping and association analyses, especially in species lacking of prior information of reference genome sequences. The most common GBS procedure employs a restriction-site-associated DNA sequencing library (Baird et al. 2008), therefore only SNPs at these targeted locations were genotyped. In addition to the limited genome coverage, GBS genotyping results can't be easily shwered by different research groups because different SNP loci were assayed in all individuals. In contrast, SNP arrays were more cost-effective when genotyping large numbers of individuals. With flexible marker selection, the dense SNP arrays can be tailored for large scale SNP genotyping on the whole genome-wide, as well as fine-scale mapping on the target region of interest.

Multiple SNP genotyping platforms have been developed for mid- to high-throughput genotyping, including Sequenom Massarray (Gabriel et al. 2009), SNPstream (Mulcahy et al. 2001), SnaPshot Multiplex (Ben-Avi et al. 2004), Illumina BeadChip (Oliphant et al. 2002), and Affymetrix GeneChip (Dalma-Weiszhausz et al. 2006). Of these platforms, Illumina and Affymetrix provide the highest density of SNP probes. The protocols of the two platforms both adopted specialized equipment to measure the signal intensity associated with the hybridization of fragmented single-stranded DNA to its corresponding nucleotide probes on the array. SNP genotype inferences were then generated by subsequent processing and analysis of probe intensity measures (LaFramboise 2009). Moreover, Illumina Infinium iSelect and Affymetrix Axiom myDesign platforms allow customized design of genotyping panel for genomic selection

and screen, which enable studies of species or populations not supported by standard products (Ragoussis 2009).

High-throughput SNP genotyping arrays have been rapidly developed and applied to achieve significant genetic enhancement in agriculture species. For instance, in cattle, the leading agriculture and commercial breeding species, SNP array technologies have been used for genetic mapping and multitude of genome-wide association studies (GWA studies) to search for polymorphisms underlying variations in complex traits (Miller 2010). The first commercially available bovine high-density SNP array was designed with Affymetrix GeneChip platform, which included 10K SNPs obtained from the Bovine Genome Sequencing Project and CSIRO (Australia's Commonwealth Scientific and Industrial Research Organization). This SNP array was used to characterize haplotype blocks and tag SNPs in 1000 Holstein-Friesian bulls (Khatkar et al. 2007). As required by high-resolution linkage and association mapping, Affymetrix released a new Bovine SNP array designed with the 10K original SNPs and additional 15K novel SNPs that obtained from the sequencing initiative. It has been applied in investigation of genome-wide breeding value prediction (Luan et al. 2009), as well as refined mapping of QTLs associated with recessive disorders (Charlier et al. 2008) and twinning-rate in cattle (Kim et al. 2009). In parallel, Illumina also launched an improved SNP array in cattle, Illumina Bovine SNP50 assay. In spite of public domain SNPs, it also includes 25,833 novel SNPs identified by deep sequencing of reduced representation libraries (RRL) from DNA of 66 cattle (Matukumalli et al. 2009). This assay was successfully used to predict genetic merit in domesticated breed strains (VanRaden et al. 2009; Wiggans et al. 2009). It was also applied to elucidate the DNA polymorphisms responsible for growth studies in crossbred

beef cattle (Snelling et al. 2010), Johne's disease QTL study in Holstein cattle (Pant et al. 2010), and other large-scale and high-impact studies (Decker et al. 2009; Hayes et al. 2009). Bovine genotyping arrays with higher-density of markers, Illumina BovineHD 777K and Affymetrix BOS1 678K array, have been released recently. The performance test of these two arrays indicated that they were well designed, with over 98% SNPs successfully genotyped (Rincon et al. 2011). With other agricultural species, high-throughput SNP arrays have also been developed for horse (McCue et al. 2012), pig (Ramos et al. 2009), sheep (Miller et al. 2011), and chicken (Groenen et al. 2011).

High-throughput SNP array genotyping technologies in aquaculture species started more recently than that of livestock and agriculture plants. The Atlantic salmon Illumina BeadArray for massively parallel genotyping of SNPs started with 16.5K putative SNPs as facilitated by the International Collaboration to Sequence the Atlantic salmon genome (ICSASG) (Kent et al. 2009). Due to the tetraploid status of the Atlantic salmon genome, only 35% of the SNPs on the array were validated in subsequent analysis (Yanez 2014). An improved SNP array with approximately 6K markers was then built for salmonids genetic mapping (Lien et al. 2011). It was also used in the analysis of genetic differences between domesticated and wild salmon (Karlsson et al. 2011), and analysis of QTLs associated with body weight (Gutierrez et al. 2012). Recently, a new generation of SNP genotyping platform in Atlantic salmon have been developed with 200K markers. The performance tests revealed 79.6% of total SNPs were validated as high quality based on clustering properties (Yanez et al. 2016). In catfish, the first SNP array was designed with Affymetrix Axiom myDesign platform, which included 250,113 SNPs derived from RNA-seq and whole genome sequencing data (Liu et al. 2014). The performance of the catfish

250K SNP array were evaluated by genotyping samples from backcross families and unrelated wild populations, resulting in a conversion rates of approximately 80%, and polymorphic rate over 50% for the examined samples (Liu et al. 2014). The catfish 250K SNP array was also used to construct a dense SNP linkage map with 576 fish from 3 mapping families (Li et al. 2015). Based on this linkage map, GWA studies by genotyping 6 families of F2 generation of channel and blue catfish hybrids identified significant QTLs associated with columnaris disease resistance (Geng et al. 2015). With other aquaculture species, medium- to high-throughput SNP genotyping arrays have also been developed in oysters (Lapegue et al. 2014), common carp (Xu et al. 2014), black tiger shrimp (Baranski et al. 2014), and rainbow trout (Palti et al. 2015).

CHAPTER 2. GENOME-WIDE SNP IDENTIFICATION AND DEVELOPMENT OF A HIGH-DENSITY SNP ARRAY

Abstract

Genome-wide molecular markers facilitate the dissection of genetic basis of quantitative traits. Single nucleotide polymorphisms (SNP) were one of the most widely applied molecular markers because of their high abundance and the adaptability for automated genotyping. In this work, we developed an ultra-high density SNP genotyping array for channel catfish (*Ictalurus punctatus*) and blue catfish (*Ictalurus furcatus*). SNP were first obtained from mining from the whole genome sequences, RNA-seq, and GBS data over 900 channel catfish and 200 blue catfish from previous experiments. Criteria-based quality filtration and Affymetrix in-silico probe converting test result in a final set of 690,662 high quality SNPs, comprising 581,002 channel catfish specific, 44,694 blue catfish specific, 19,124 inter-species, and 45,842 ubiquitous SNPs. A total of 48,434 strain-specific markers for 5 domestic and wild strains were also included on the array, which should be useful in study channel catfish domestication events. The selected SNPs were relatively evenly distributed on the genome reference, with 0.8 kb interquartile range of intervals. They cover 98.6% of the genome scaffolds and 93.9% of the BAC-based physical map contigs. The high coverage benefits the detection of linkage disequilibrium in further genome-wide association studies and fine-scale linkage mapping. To well represent gene-associated regions, 206,580 gene-associated SNPs were from 25,188 genes annotated from the channel catfish reference genome sequence, or from the transcripts assembled from RNA-Seq datasets. 2,000 additional probes generated from non-polymorphic genomic regions were also introduced as negative controls. This is the highest density SNP array

developed in aquaculture species, which provide valuable resources for further construction of genetic map, genome-wide association studies, and selective breeding.

Introduction

Catfish is the primary aquaculture species in the US, accounting for approximately 60% of US aquaculture production. However, the catfish industry is facing great challenges including escalating feed and fuel costs, intense international competition, and devastating diseases. Many economically important traits, including disease resistance, hypoxic oxygen adaption, growth rate, etc., are most often controlled by a group of genes involving a complex cascade of regulation. Application of genetic markers facilitate identification of QTLs that significantly associated with traits of interest and dissection of the genetic basis, providing benefits of the intensity and accuracy in selective breeding.

Single nucleotide polymorphisms (SNPs) are now the choice of genetic markers as a consequence of the advancement of sequencing power and improved computational capacities in analyzing single nucleotide variations on a genome-wide scale. Most often, SNPs are bi-allelic that is amenable to automated genotyping (Fahrenkrug et al. 2002). Compared with other genetic markers, SNPs have the most abundant resources across the genome (Shastry 2002), which makes it possible to coverage most of genes that are in linkage disequilibrium with the markers and are efficient for genome-wide association studies (GWAS). For instance, SNPs based GWAS analysis for performance and production traits in chicken (Wang et al. 2016), pig (Bergfelder-Druing et al. 2015), cattle (Crispim et al. 2015), horse (Dupuis et al. 2011), and sheep (Goher et al. 2010).

Until recently, whole genome-scale SNP identification was a great challenge for most of the non-model species. The increasing sequencing depth and accumulating of NGS datasets enabled the accurate identification of SNPs from genomes of various aquaculture organisms, including salmon, rainbow trout, and catfish. With the availability of enormous number of SNPs, the next challenge is how to apply them in QTL mapping and GWAS efficiently and economically.

Among many genome resources, high-density SNP arrays were the most important because such arrays allow genotyping to be conducted efficiently and automatically. A variety of SNP genotyping platforms have been developed to deal with different marker quantities and sample sizes. As for genotyping of less than hundreds of SNPs, Taqman, SNPlex, Sequenom MassArray (Sequenom, San Diego, CA), and other electrophoresis or fluorescence-based homogenous assays could be used. These platforms were economical and flexible in selection of markers. To investigate more than thousands of markers, array technologies were more efficient and economical than other methods, such as sequencing. Illumina iSelect HD Custom BeadChip (Illumina, San Diego, CA) and Affymetrix GeneChip Custom Array (Affymetrix, Santa Clara, CA) were two the most widely used platforms. Customized probes can be included on the array to inspect specific SNPs of interest. More recently, Affymetrix adopted the Axiom genotyping technology that allows investigation up to 2.6 million SNPs or CNVs simultaneously (Hoffmann et al. 2011). Up to date, Axiom genotyping platforms have been designed in cattle, mouse, chicken, catfish, Atlantic salmon, and rainbow trout. With the available of RNA target site genotyping solution, variants in microRNA could also be investigate, making it ideal for supplementing GWAS or gene regulation studies (Schmit et al. 2015). In spite of the low

unit cost per genotype (~0.03 per SNP/genotype), the total costs for the high density SNP arrays with enormous number of SNPs may not be affordable for researches working with “minor” species.

In catfish, the first SNP array was developed with previously identified 250,113 SNPs (Liu et al. 2014). This array has been very useful for various genetic analysis including genetic linkage mapping (Li et al. 2015; Liu et al. 2016a), and QTL mapping or GWAS (Geng et al. 2015). The advancement of array technologies enables fabrication of arrays with expended capacity at a similar cost. With the increasing sequencing depth and the guidance of reference genome sequence of catfish, a total of 690,662 SNPs were used for construction on the second generation of catfish SNP array, which should have a better genome coverage and enhanced marker densities, useful for construction of genetic maps, GWAS, and QTL analysis.

Materials and methods

SNP identification

In order to identify SNPs from the whole genome, Illumina sequencing data from various studies involving fish with diverse genetic background were collected. For channel catfish, a total of 2,696.5 million reads from RNA-seq of 815 fish and 2,400 million reads from whole genome sequencing of 150 fish were collected for SNP identification, with an average genome coverage of x500 (Table 1). For blue catfish, data used for SNP identification included over 428 million reads from GBS data of 190 individuals and 477.5 million reads from RNA-seq data of 49 individuals.

Table 1. Sample and data size for development of 690K SNP array.

Project	Species	Sample size	Data type	Data accession	Reads (x10 ⁶)
Gene-associated SNP mining	Channel, blue catfish	66	RNA-seq	SRA025099	506
Studies on catfish domestication	Channel catfish	150	WGS	SRX275877	2,400
Gonad development	Channel catfish	279	RNA-seq	SRP018265, SRP067841, Unpublished	1,185
Barbel regeneration	Channel catfish	90	RNA-seq	Unpublished	395
Skin regeneration	Channel catfish	15	RNA-seq	Unpublished	197
ESC challenge	Channel catfish	180	RNA-seq	SRP009069	197
Columnaris challenge	Channel catfish	204	RNA-seq	SRP017689, Unpublished	500
Feed deprivation	Blue catfish	30	RNA-seq	SRP020252	194
SNP mining with GBS	Blue catfish	190	GBS	SRP039640	428
Total	-	1,204	-	-	6,002

To reduce sequencing artifacts and improve the SNP quality, raw sequencing reads from all the studies were first subjected to quality control with Trimmomatic (version 5.5.2). Adaptor sequences, ambiguous nucleotides (N's), extreme short reads (< 25 bp) were removed. Low quality bases were identified and trimmed with a sliding window method, bases within a window size of 4 was cut once the average quality is less than 15. For reads generated from GBS, chimeric sequences were eliminated by trimming the sequence at the corresponding restriction enzyme site. The clean reads generated by WGS and GBS were then aligned to the reference genome sequence (Liu et al., 2016) with BWA-MEM (version 0.7.12). To acquire high sensitivity and accuracy for RNA reads alignment, a 2-pass

alignment method is used by STAR aligner (version 2.4.0j). The results of the alignment were exported in BAM format for subsequent analysis (Li et al. 2009).

Prior to SNP identification, the BAM files were processed with Picard tools (version 1.119) to identify and remove redundant copies of duplicates. BAM files of WGS reads were subjected to local realignment of regions near INDELs and recalibration of base quality scores with GATK (version 3.3) to improve the accuracy of variant calling. For BAM files of RNA-seq reads, “SplitNCigarReads” commands of GATK is used to split reads which span multiple exons and to trim overhangs. Ambiguously mapped reads were removed using SAMtools (version 0.1.19) under the criteria of a minimum MAPQ score of 20. For paired-end reads, only those mapped in a proper pair were kept for further analysis. After that, all of the alignment files were piled up together for variant calling using Varscan (version 2.3.7). The putative SNPs were called with the thresholds of minor allele frequency greater than 0.05, minimum read base quality of 20, strand-filter less than 90%, and minimum read depth of 10. Sequences of 71-bp spanning each SNP were extracted, with 35-bp upstream and 35-bp downstream of the SNP base, respectively. To avoid false positive SNPs caused by ambiguous mapping of duplicative regions, the 71-bp fragments were aligned to the reference genome sequence with BLAST suite Blast + (version 2.2.29). SNPs with flanking regions that mapped to multiple sites or low complexity and repetitive regions on the reference genome sequence were filtered.

SNP selection and array design

In order to provide a high level of flexibility, we chose to use the Affymetrix Axiom myDesign GW genotyping platform to develop the new SNP array. Axiom technology

adopt ligation assay-based genotyping platform: the genomic DNA were firstly amplified and randomly fragmented into 25 to 125 bp fragments. Then these fragments were purified and hybridized to the glass probes on the custom array plates. Non-specific hybridization fragments were washed to avoid background noise. The probe-target complexes were then ligated specifically to solution probes that bearing attachment sites for one of two dyes. A/T and C/G SNPs require additional interrogations, because the two alleles match the same dye and therefore, an extra probe sets on the array were required to distinguish them. Following ligation, the genotyping array is stained and imaged on GeneTitan MC Instrument (Axiom 2.0 Assay Manual Workflow).

SNP selection was performed in multiple steps using different criteria regarding different SNP types. All the filtration parameters were set with the aim of a well spacing coverage on the whole genome and removal of false positive sites. All the original SNPs were classified into different groups and selected in a certain order: SNPs from gene coding regions and UTR regions were first selected, then SNPs from non-coding regions were added, finally, species-specific SNPs and strain-specific SNPs were added to the pool of candidate SNPs. Custom-made scripts were used to perform the selection according to the following criteria: (1) To avoid non-specific hybridization, the flanking sequences of selected SNPs should not have other SNPs or simple nucleotide repeats; (2) For practical application in SNP genotyping assays, only bi-allelic SNPs were selected; (3) To acquire high-polymorphic rate, SNPs with a minor allele frequency greater than 0.1 were preferentially selected; (4) To acquire a high-capacity, A/T or C/G SNPs were not selected unless absolutely needed; (5) To minimize the effects of GC content on the signal intensity, the GC percentage of the SNPs flanking sequences should be between 30-70%. For SNPs

in the gene-coding regions, once a new SNP was included into the candidate pool, SNPs from the flanking regions of 200 bp were not added any more. For SNPs in the non-coding regions, once a new SNP was included into the candidate pool, SNPs from the flanking regions of 350 bp were not added. SNPs from unmapped scaffold and contigs apart from reference genome sequences were also included regardless of their distances.

All catfish SNPs that passed this filter were submitted to Affymetrix Bioinformatics Services for *in-silico* probe converting test, where the performing quality of the SNPs were evaluated. Upstream and downstream probes flanking the SNPs were assigned with a p-convert value (0.0 to 1.0), respectively. Probes with high p-convert values were more likely to be successful genotyped. A p-convert value threshold was determined by excluding the tail of lowest performing probes to facilitate selection of final SNP list. Markers with at least one probe that passed the p-convert value threshold were retained. For SNP Markers with both of the two probes that pass the p-convert value threshold, the probe with greater p-convert value was selected. For the SNPs with only probes of low p-convert values, both the two probes were included to cover the genome region.

In addition to the polymorphic SNPs, 2,000 probes generated from non-polymorphic genomic regions were also introduced as data quality control (DQC) probes. Of which, 1,000 probes were selected with A or T at the 31st base, and 1,000 DQC probes were selected with G or C at the 31st base. The DQC probes along with the final list of SNPs were submitted to Affymetrix for fabrication of Axiom GW genotyping array.

SNP array performance evaluation

Fish sources and DNA isolation

A total of 480 catfish were genotyped to assess the performance of SNP array, including 396 channel catfish of delta select line which provided by USDA-ARS Warmwater Aquaculture Research Unit, and 84 catfish hybrids generated by backcrossing the interspecific hybrids (channel x blue) with male channel catfish.

DNA samples were prepared following the procedures as previous described (Waldbieser et al. 2001). In brief, the fish were euthanized with tricaine methanesulfonate (MS 222) at 300 mg/l before blood collection. For each individual, 500 μ l blood samples were collected with a 1 ml syringe and immediately expelled into a 50-ml tube with 20 ml of DNA extraction buffer (10 mM Tris, 100 mM NaCl, 25 mM EDTA, 0.5% SDS, 0.1 mg/ml proteinase K, pH 8.0). The extracted blood samples were then incubated at 55° C overnight and then extracted with phenol and chloroform. Following extraction, DNA samples were precipitated by isopropanol and collected by brief centrifugation, washed with 70% ethanol and resuspended in reduced EDTA TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). DNA samples were quantified using spectroscopy by Nanodrop (Thermo Scientific) and checked for integrity by 1% agarose gel electrophoresis stained with ethidium bromide.

SNP genotyping

DNA samples were diluted to 50 ng/ μ l and genotyped with the catfish 690K SNP array at GeneSeek (Lincoln, Nebraska, USA). The signal intensity data of each probe on the array were stored in CEL files, which were analyzed by Axiom Analysis Suite (version

1.1.0.616) for quality control and genotype calling. Samples with a Dish value greater than 0.85 and SNP call rates greater than 95% were retained for subsequent analysis.

Following the genotyping step, SNPolisher (an R package developed by Affymetrix) is used to generate quality metrics and classify all the SNPs into six types. “PolyHighResolution” indicate that both of the two alleles of a SNP were detected. The signal data of all the samples also formed into three distinct clusters with good resolution; “NoMinorHom” indicate SNPs with two clusters of signal data, with no example of the minor homozygous genotypes; “MonoHighResolution” include SNPs with only one clusters identified. “OTV” refers to off-target variants, indicating SNPs with an OTV cluster that caused by sequence dissimilarity between probes and target genome regions (Didion et al. 2012). “CallRateBelowThreshold” were the SNPs with call rates below threshold, but other cluster properties were above threshold. “Other” were the SNPs with one or more cluster properties below the threshold. In most cases, SNPs classified as “PolyHighResolution”, “NoMinorHom”, “MonoHighResolution”, were considered as convertible SNPs. SNPs classified as “PolyHighResolution” and “NominorHom” were considered as polymorphic SNPs.

Results and discussion

SNP identification and selection

With over 6,002 million reads from 1,204 catfish, a total of over 9.6 million putative SNPs were identified, including 8.6 million SNPs from channel catfish and 3.8 million SNPs from blue catfish. As channel catfish is assumed to have undergone the teleost-specific genome duplication (TSGD) (Liu et al. 2016b). A considerable number of

gene sets (over 3,000) were identified to contain two or more copies (Liu et al. 2016b). To remove false SNPs potentially caused by duplicate copies (paralogs) of genome regions, all the SNPs with 35 bp up- and down- stream flanking sequences (71 bp in total) were mapped onto the latest reference genome assembly using BLAST. ~3 million SNPs from channel catfish and ~0.7 million SNPs from blue catfish can be aligned to more than one site of the reference genome sequences (Table 2). These sequences were filtered. The 35 bp flanking regions of SNPs were used to develop the array probes. There should be no sequence variants in such a region. Therefore, any SNPs with an interval shorter than 35 bp were also filtered. As revealed by Table 2, over 3.3 million SNPs from channel catfish and 2.7 million SNPs from blue catfish were removed in this step, indicating that the overall distribution of SNPs was uneven across the genome. To ensure the probe specificity, SNPs near simple nucleotide repeats were also removed, ~15,000 SNPs from channel catfish and ~8,000 SNPs from blue catfish were filtered in this step. As the Axiom genotyping platform can only discriminate bi-allelic SNPs, over 340,000 SNPs from channel catfish and 38,000 SNPs from blue catfish were removed because there were more than two alleles for these SNPs.

A total of 2.4 million SNPs initially passed in house selection as described above. These SNPs were submitted to Affymetrix Bioinformatics Servers for *in-silico* analysis to assess the predicted performance of the SNP probes on Axiom arrays. For each SNP, both the forward and reverse probe were evaluated with a random forest model to predict the probability that the SNP will be convertible on the array. The model considers multiple factors including probe sequence content, binding energy, as well as the expected degree of specificity of hybridization (personal communication with Lucy Reynolds,

Bioinformatics Services Engineer). SNP probes with high p-conver values were expected to convert on the array. About 1.8 million SNPs from channel catfish and 0.3 million SNPs from blue catfish were with a p-conver value greater than the threshold of 0.5.

Table 2. Summary of SNPs identified in channel and blue catfish.

	Channel catfish	Blue catfish	Total
Number of putative SNPs	8,644,975	3,850,662	9,662,958
Number of unique mapping SNPs	5,667,615	3,138,197	6,535,576
Number of SNPs without adjacent SNPs in 35 bp	2,270,690	347,958	2,512,422
Number of SNPs passed flanking sequence complexity inspection	2,199,952	340,921	2,435,602
Number of bi-allelic SNPs	2,184,946	332,934	2,418,168
Number of SNPs passed Affymetrix <i>in-silico</i> test	1,839,478	294,593	2,046,175

SNPs from four channel catfish domestic strains and a wild population were also selected with the same procedure (Table 3). After filtration, about 1 million qualified SNPs were retained for each strain/population.

Table 3. Summary of SNPs identified in channel catfish domestic strains and wild population.

	Thompson	Hatchery	Marion	USDA103	Wild
Number of putative SNPs	5,723,886	4,270,527	4,817,089	4,900,634	5,488,642
Number of unique mapping SNPs	3,992,954	2,861,190	3,283,718	3,345,046	3,852,203
Number of SNPs without adjacent SNPs in 35 bp	1,487,079	1,032,160	1,184,548	1,225,949	1,442,459
Number of SNPs passed flanking sequence complexity inspection	1,455,201	1,009,026	1,155,782	1,196,716	1,414,789
Number of bi-allelic SNPs	1,443,954	1,000,560	1,146,365	1,187,187	1,403,727
Number of SNPs passed Affymetrix <i>in-silico</i> test	1,224,483	846,622	969,303	1,004,560	1,191,020

As there were many more than needed SNPs passed the p-convert evaluation, additional criteria were applied to rank the order, including: (1) SNPs with p-convert value greater than 0.65; (2) SNPs with MAF greater than 0.1; (3) SNPs were not A/T or C/G; (2) SNPs with GC percentage of flanking sequences between 30-70%.

SNPs included on the 690K array

The high-density new SNP array provided better coverage of gene and reference genome sequences. With the enlarged sample and data size, more high quality SNPs could be identified on gene associated regions and previously unmapped scaffolds. The final

SNPs included on the new catfish SNP array were summarized in Table 4. A total of 690,662 SNPs were included, including 206,580 gene-associated SNPs and 484,082 anonymous SNPs. Of the gene associated SNPs, 206,270 were from 24,878 genes annotated from the channel catfish reference genome sequence, the other 310 SNPs were from 521 genes annotated from transcript sequences via *de novo* transcriptome assembly. Inclusion of gene associated SNPs could enhance the conversion rate as genes and their associated sequences should be more unique in the genome than the non-coding regions of the genome. As genes were not entirely evenly distributed, inter-marker spacing is not equal. The 484,082 SNPs from non-coding regions would fill the gaps.

Table 4. Summary of the catfish 690K SNP array.

SNP array	Number
Number of gene-associated SNPs	206,580
Number of SNPs from anonymous genome regions	484,082
Total number of SNPs on array	690,662
Number of SNPs tiled with single probe	687,757
Number of SNPs tiled with two probes	2,905
Total number of probes	693,567

To ensure the performance, markers with high quality probes and MAFs were included on the array. A total of 693,567 SNP probes were synthesized for interrogation of these 690,662 SNPs. Most of the probes were with a high p-conver value (greater than 0.65) (Figure 1). Only 2,905 SNPs of which were tiled with two probes. In addition to SNP

probes, 2,000 DQC probes were included on the SNP array serving as negative controls. As for linkage mapping, GWAS, and other genetic analysis, common variants were more important compared with rare variants, therefore, most of the included SNPs were common variants with a MAF greater than 0.1 (Figure 2).

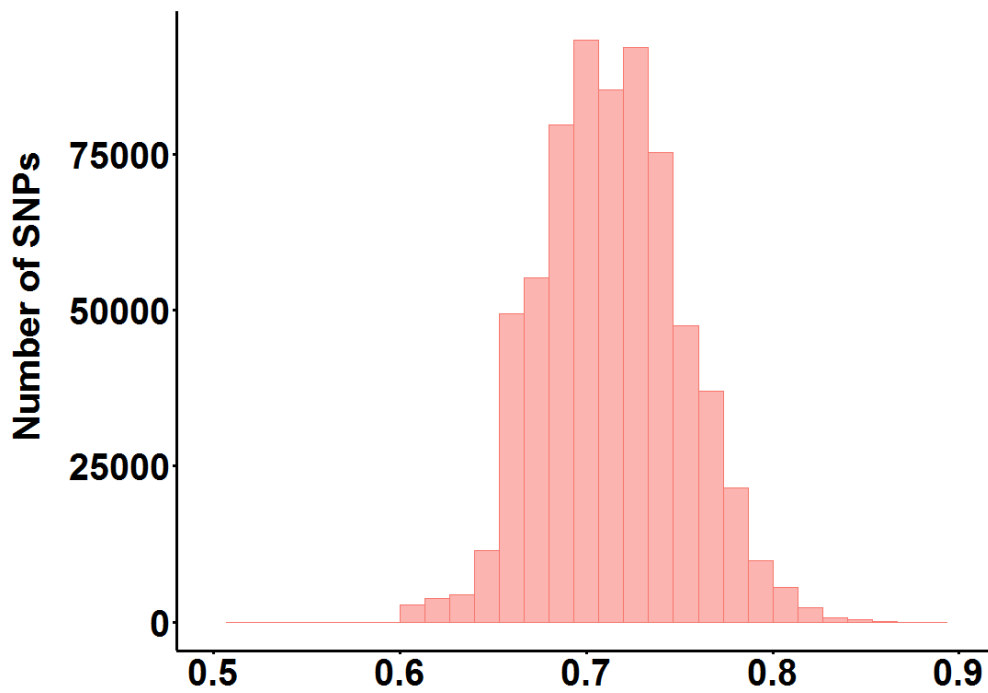


Figure 1. Distribution of SNP probes based on p-convert values.

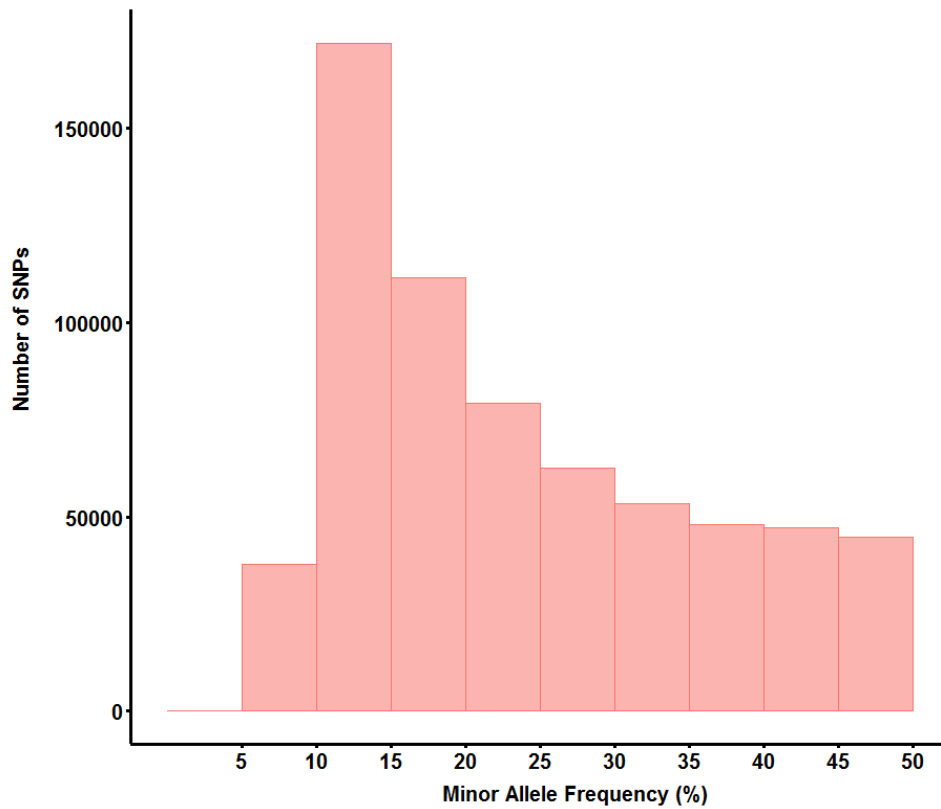


Figure 2. Distribution of SNPs based on MAF.

The 690K catfish SNP array also included species and strain specific SNPs (Table 5), which should be useful for genetic analysis of the interspecific hybrid system and intraspecific crossbreeding. Of the total SNPs on the array, 581,002 SNPs were specifically from channel catfish, 44,694 were exclusively observed in blue catfish, 19,124 were interspecific SNPs identified between channel catfish and blue catfish, and 45,842 were identified both in channel and blue catfish. To facilitate catfish domestication studies, 48,434 SNPs from four catfish aquaculture strains that originate from different geographic locations were also included on the array. These four strains possess different production traits such as growth rates, disease resistance, and adaptation to environmental stresses. A total of 6,622 SNPs were specifically from a wild channel catfish population of Coosa River,

Alabama, which should be useful for genetic analysis of genomic regions with selective signatures.

Table 5. Summary of the species/strain-specific SNPs on catfish 690K SNP array.

SNP array	Number
Channel catfish-specific SNPs	581,002
Blue catfish-specific SNPs	44,694
Inter-species SNPs	19,124
Channel-blue both possessed	45,842
Strain-specific SNPs	48,434
Domesticate strain (Thompson)	12,672
Domesticate strain (Hatchery)	9,498
Domesticate strain (Marion)	10,309
Domesticate strain (USDA 103)	9,333
Wild population	6,622
Total	690,662

Distribution of SNP spacing

One of the most important goals of the SNP array development is to have a good coverage of the genome with relatively even distribution across the entire genome. The locus of SNPs on the reference genome sequences were set as the coordinates to assess their overall distribution. A total of 659,912 (95.5%) SNPs were developed from the reference genome reference, which span a total of 778 Mb, approximately 99.1% of the reference genome sequence. As shown in Figure 1, almost all of the regions were covered

by SNPs, except a few highly repetitive regions from which convertible SNP probes cannot be designed. A total of 30,750 (4.5%) SNPs were developed from other genomic resources, including 15,108 SNPs from scaffolds that have not been anchored to the reference, 521 SNPs from transcript sequences via *de novo* transcriptome assembly, and 15,121 SNPs from bacterial artificial chromosome (BAC) end sequences (BES) constructed contigs which were not able to integrate with linkage map. Taking together, the total SNPs covered 98.6% of the genome scaffolds and 93.9% of the BEC based physical map contigs.

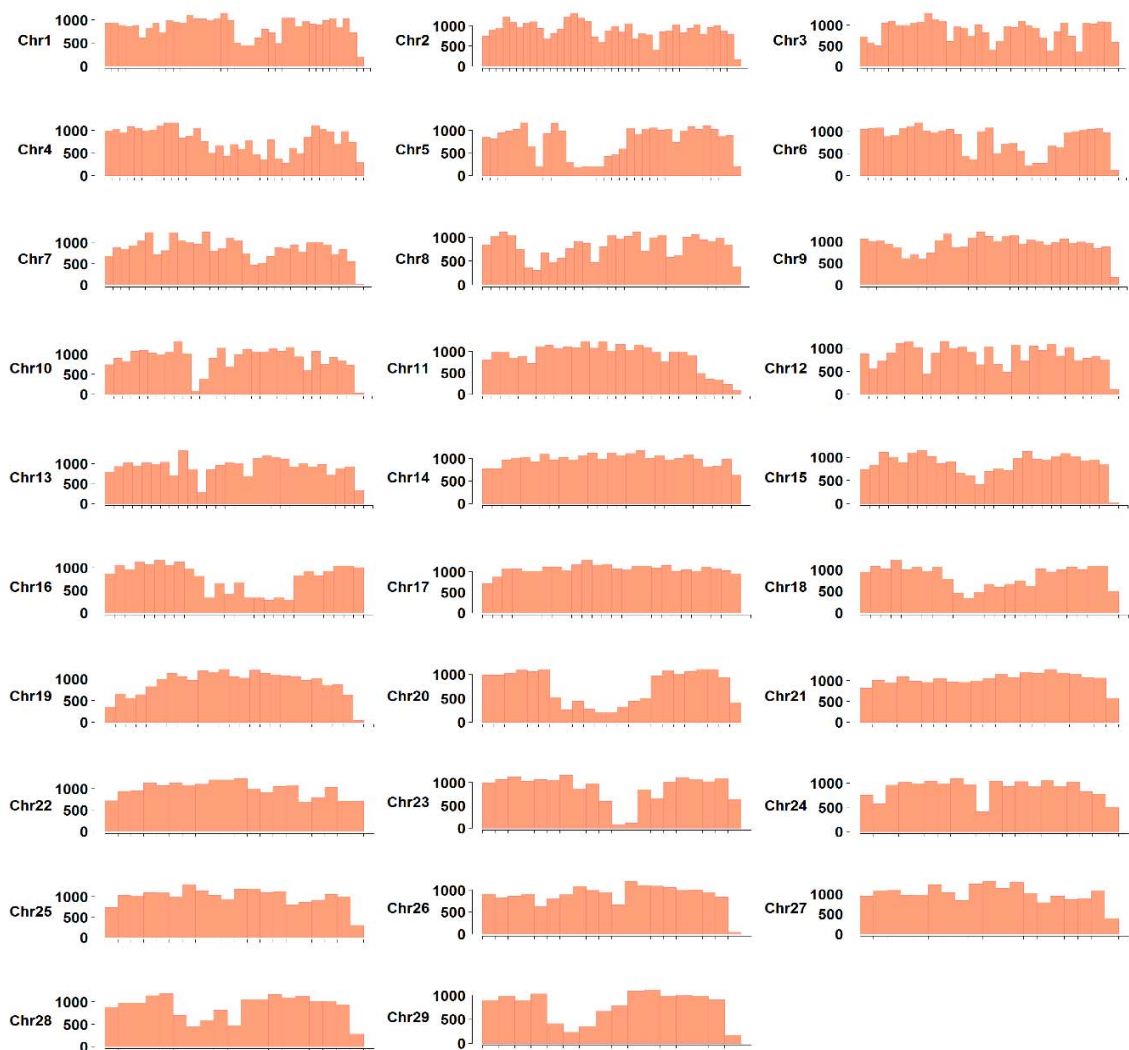


Figure 3. Genome distribution of SNPs on the 690K SNP array.

The inter-SNP spacing were evaluated using their physical position on the reference genome or other genomic contigs. As shown in Figure 4, most of the SNPs had an inter-SNP spacing less than 2,500 bp. Of which, approximately 45,000 SNPs had a small inter-SNP spacing of less than 200 bp, ~100,000 SNPs had an inter-SNP spacing of 200-400 bp, the largest number of SNPs (~110,000) had an inter-SNP spacing of 400-600 bp. A total of ~85,000 SNPs had a marker spacing of 600-800 bp, ~60,000 SNPs had a marker spacing of 800-1,000 bp. ~100,000 SNPs had a marker spacing of 1,000-1,500 bp, ~55,000 SNPs had a marker spacing of 1,500-2,000 bp. A relative small part of SNPs had a marker spacing greater than 2,000 bp. As shown in Figure 3, the markers intervals have an interquartile range of 500-1,300 bp.

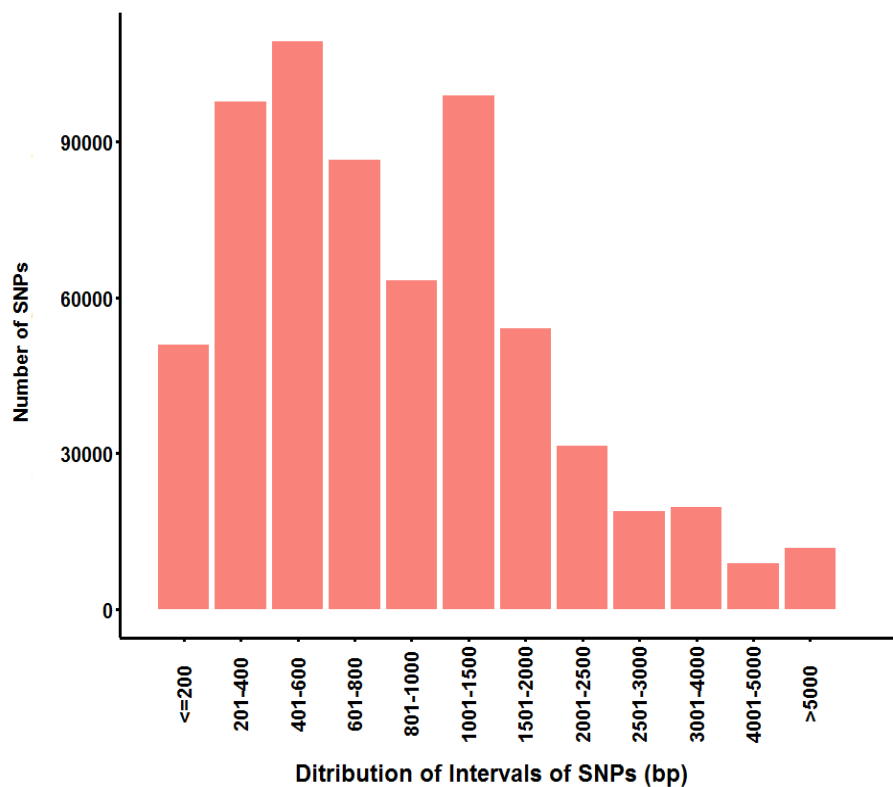


Figure 4. Distribution of inter-SNPs spacing for SNPs included on the array.

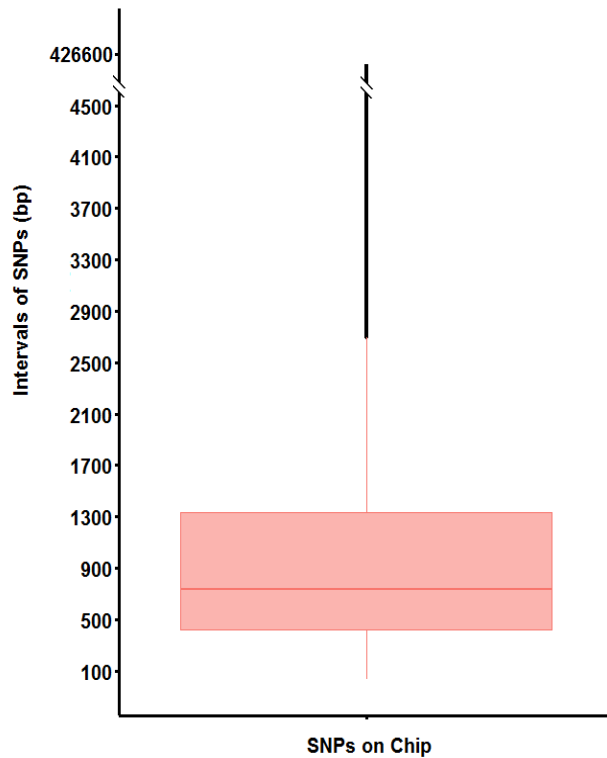


Figure 5. Box-plot of inter-SNPs spacing for SNPs included on the array.

Performance of the catfish 690K SNP array

Genotyping performance of the SNP array

Performance of the SNP array was examined by genotyping catfish DNA samples from hybrid backcross families and channel catfish domesticated families. As summarized in Table 6, 473 of 480 catfish samples (98.5%) were successfully genotyped after sample quality control. In backcross hybrids samples, a total of 597,323 (86.1%) SNPs were converted, and 504,265 (72.7%) were polymorphic in these samples. The average call rate of DQC qualified samples was greater than 99.2%. In channel catfish samples, a total of 578,868 (83.5%) SNPs were converted, of which 467,821 (67.5%) were polymorphic.

Despite more channel catfish samples were processed, much higher percentages of SNPs were converted and polymorphic in backcross hybrids samples. This may be caused by the interspecific and blue-specific SNPs included on the array. The backcross hybrids possess fraction of “blue catfish” genome materials, therefore, 35,156 blue-specific SNPs and 15,538 interspecific SNPs were polymorphic in these samples.

Table 6. SNP metrics summary.

	Backcross hybrids	Channel catfish	Total
Samples processed	84	396	480
Samples passed QC	81 (96.4%)	392 (98.9%)	473 (98.5%)
PolyHighResolution	292,185 (42.1%)	326,411 (47.1%)	401,815 (57.9%)
NoMinorHom	211,980 (30.6%)	141,410 (20.4%)	133,803 (19.3%)
Total Polymorphic SNPs	504,265 (72.7%)	467,821 (67.5%)	535,618 (77.2%)
MonoHighResolution	93,058 (13.4%)	111,047 (16.0%)	62,262 (8.9%)
Total converted SNPs	597,323 (86.1%)	578,868 (83.5%)	597,880 (86.2%)

Conclusion

In this study, we developed the catfish 690K SNP array using Affymetrix Axiom technology. The final set SNPs on the array comprising channel catfish specific, blue

catfish specific, and inter-species SNPs. A subset of strain-specific markers from domestic and wild channel catfish populations were also included on the array, which should be useful in study channel catfish domestication events. The selected SNPs were relatively evenly distributed on the reference genome sequence, and provided a high coverage on the whole genome, which should benefit the detection of linkage disequilibrium in further genome-wide association studies and fine-scale linkage mapping. Distribution of minor allele frequency indicated that most of the SNPs included on the array were common variants (vast majority of SNPs have a minor allele frequency >0.1). The evaluation of SNP array performance by genotyping samples from catfish backcross hybrid families and channel catfish families suggested high polymorphic rates (~77%) and high SNP conversion rates (~86%). The catfish 690K SNP array should serve as valuable resources for both industry and research such as genetic mapping and analysis of quantitative traits.

CHAPTER 3. CONSTRUCTION OF A HIGH-DENSITY GENETIC MAP AND ITS INTEGRATION WITH GENOMIC SEQUENCES OF CHANNEL CATFISH

Abstract

Catfish is the leading aquaculture species in United States of American. However, the catfish industry is facing great challenges due to international competition, escalating feed and energy cost, and devastating diseases. Genetic and genomic studies can greatly contribute to improve the economical traits and performance of domestic species, providing enhanced production efficiency. Construction of accurate and high density genetic linkage map is fundamental for genetic and genomic studies. Great advances in sequencing technologies and genotyping platforms enabled the construction of high-density and accurate genetic linkage maps, especially for non-model organisms with relative scarce genome resources. In the present work, I constructed a new generation of channel catfish genetic linkage map with improved marker density and resolution with four large reference families genotyped using the catfish 690K SNP array. Over 250,000 SNPs were finally placed on the linkage map, which contains, to the best of our knowledge, the highest marker density among all the non-model organisms. The total estimated genetic size of the new linkage map is 3,004

cM with a resolution of 0.1 cM for sex-averaged genetic map. The sex-specific linkage maps constructed in each sex revealed that the recombination rate is 1.4-fold larger in females than in males. After integration with the recently constructed catfish reference genome assembly, 1,546 of the whole-genome sequence contigs were anchored to the linkage groups, covering a physical length of 766 Mb that account for about 97.8% of the 783 Mb genome sequence assembly. Additional 950 previously unmapped scaffolds were

placed onto the linkage map, which greatly improves the catfish reference genome assembly. With the unprecedented coverage, it should also serve as a valuable tool for genetic analysis, especially GWAS and fine-scale QTL mapping for genes associated with economically important traits.

Introduction

Genetic maps are essential tools for genomic and genetic studies. It provides a frame of the genome structure with landmarks of DNA markers along chromosomes; it also provides the inference of how the rate of recombination varies across the genome (Dumont et al, 2008). Recombination is an important sequence feature, which serve as a critical parameter in governing the degree and nature of intraspecific diversity as well as interspecific divergence. Revealed by Roselius et al., the recombination rate is expected to significantly correlate with local levels of nucleotide content, and the rate of adaptive evolution (Roselius et al. 2005). There is also an increasing awareness that the local recombination rate regulates the evolution of nucleotide composition by GC-biased gene conversion (Weber et al. 2014). As regulated by recombination rate, genomic landscape played critical role in species differentiation (Burri et al. 2015).

Another important usage of genetic map is to anchor and order scaffolds along chromosomes (Beldade et al. 2009). High-density linkage map is a valuable tool to make full use of the unprecedented power provided by NGS technology. Despite a few long reads sequencing technologies which were not widely used until very recently, most of the sequencing data generated by NGS were relatively short in size. Contigs assembled through short reads were difficult to be organized on the chromosomal level. Integration of genetic

map with physical map is useful for understanding genomes from different dimensions and is essential for genomic studies, including comparative genome analysis, GWAS, and QTL mapping.

Notable progress has been accomplished in the past decades in the construction of linkage maps as a consequence of the rapid development of molecular markers. Until recently, linkage maps for more than 40 aquaculture species have been constructed using various genetic markers (Yue 2014), including AFLP, microsatellite, as well as SNPs. The density of genetic maps have increased significantly in many important teleost species such as catfish (Li et al. 2015), Atlantic salmon (Gonen et al. 2014; Lien et al. 2011), rainbow trout (Guyomard et al. 2012; Rexroad et al. 2008), Nile tilapia (Lee et al. 2005), Japanese flounder (Castano-Sanchez et al. 2010), Asian seabass (Wang et al. 2011), sea bream (Tsigenopoulos et al. 2014), bighead carp (Zhu et al. 2014), yellowtail (Ohara et al. 2005), Pacific oyster (Hedgecock et al. 2015), South African abalone (Rhode et al. 2012), and many others. High quality genetic linkage maps can provide framework for QTL localization, and facilitate marker-assisted selection and breeding in many agriculture and aquaculture species. For instance, growth-related traits have been mapped and characterized in many teleost fishes, such as Asian seabass (Wang et al. 2015a), Atlantic salmon (Wang et al. 2015a), common carp (Laghari et al. 2015), and rainbow trout (Wringe et al. 2010). Diseases resistance related traits have been mapped in Japanese flounder and applied to marker assisted selection (Yu et al. 2016). Sex markers have been identified in the genetic map of tilapia (Shirak et al. 2006), halibut (Palaiokostas et al. 2013), catfish (Ninwichian et al., 2012), and smooth half tongue sole (Song et al. 2012).

As catfish is the most important aquaculture species, much efforts on genomic and genetic studies have been conducted. In the past decades, a number of linkage maps of catfish have been constructed with different types of molecular markers, using various resource families, and with various levels of marker densities (Kucuktas et al. 2009; Liu et al. 2003; Waldbieser et al. 2001). The first catfish linkage map was constructed with microsatellite markers and two channel catfish intraspecific reference families. The microsatellite-based linkage map grouped 262 loci into 32 multipoint linkage groups which was spanned a total genetic length of 1,958 cM. The overall recombination rate in female is 3.18-fold larger than male (Waldbieser et al., 2001). In 2003, a second genetic map in catfish was constructed using an interspecific hybrid reference family with AFLP markers. The reference family was generated by backcrossing a F1 hybrid male (channel catfish female x blue catfish male) with a channel catfish female. The AFLP based genetic map is composed of 419 markers in 44 linkage groups with a genetic length of 1,593 cM (Liu et al., 2003). The inflated linkage group number indicated that more markers were need to fill the gaps and link the fragment linkage clusters. With the generation of EST sequences, more available markers were obtained. Therefore, a third catfish linkage map was constructed with 259 microsatellites and 72 SNPs. These markers were assigned into 29 linkage group and covered a total genetic length of 1,811cM (Kucuktas et al., 2009). In 2012, more genomic markers were developed from the BES-based physical contigs. Markers used to construct the fourth catfish linkage map that included 2,099 BES-based microsatellites, 235 microsatellites from EST, 127 anonymous microsatellites, 17 microsatellites from NGS data, and 79 SNPs. The total genetic length increased to 2,550cM, with an average marker interval of 1.4 cM (Ninwichian et al. 2012). Great improvement of

the linkage map was achieved with the available of catfish 250K SNP array (Liu et al., 2014). A SNP-based linkage map was constructed in channel catfish by using three full-sib reference families. A total of 54,342 SNPs were assigned into 29 linkage groups. The sex-averaged genetic map spanned 3,505 cM, with a resolution of 0.22 cM (Li et al., 2015). However, about 11% of physical map contigs were still not anchored onto the linkage map and about 25 Mb of the assembled genome sequences were not mapped. A linkage map with higher density and resolution is still desired in order to provide a high level of integration of the linkage map with the reference genome sequence. However, it is a great challenge to genotype larger numbers of markers in large reference families. The recently developed catfish 690K SNP array provided an affordable tool for the ultra-high density linkage map construction. The 690,662 SNPs on the genotyping array provided an unprecedented coverage of channel catfish whole genome, with an average interval of 1.5kb.

In this study, we constructed an ultra-high resolution genetic linkage map with 254,825 SNP markers, which is the highest density genetic linkage map among all aquaculture species. This work should provide a frame work for improving channel catfish reference genome sequence, as well as a powerful tool for QTL mapping and GWAS analysis for economically important traits.

Materials and methods

Resource families and DNA preparation

A total of 478 Delta select channel catfish from four full-sib families were used for linkage mapping. All fish were provided by USDA-ARS Warm Water Aquaculture

Research Unit. DNA samples were prepared following the procedures as previously described (Waldbieser et al. 2001). In brief, the fish were euthanized with tricaine methanesulfonate (MS 222) at 300 mg/l before blood collection. For each individual, 500 μ l blood samples were collected with a 1 ml syringe and immediately expelled into a 50 ml tube with 20 ml of DNA extraction buffer (10 mM Tris, 100 mM NaCl, 25 mM EDTA, 0.5% SDS, 0.1 mg/ml proteinase K, pH 8.0). The extracted blood samples were then incubated at 55 °C overnight and then extracted with phenol and chloroform to remove protein and DNase. Following extraction, DNA of each sample were precipitated by isopropanol and collected by brief centrifugation, washed with 70% ethanol and resuspended in reduced EDTA TE buffer (10 mM Tris-HCl, 0.1 mM EDTA, pH 8.0). The extracted DNA samples were quantified using spectroscopy by Nanodrop (Thermo Scientific) and checked for integrity by 1% agarose gel electrophoresis stained with ethidium bromide.

SNP genotyping and pedigree testing

DNA samples were diluted to 50 ng/ μ l and genotyped with the catfish 690K SNP array at GeneSeek (Lincoln, Nebraska, USA). The intensity data of each probe on the array were stored as CEL files, which were analyzed by Axiom Analysis Suite (version 1.1.0.616) for quality control. Samples with a DQC value greater than 0.85 and SNP call rates greater than 95% were retained for subsequent analysis. Genotype calling were achieved by Affymetrix AxiomGT1 algorithm. SNPolisher (an R package developed by Affymetrix) is used to generate quality metrics and classify all the SNPs into six types. SNPs classified as “PolyHighResolution” and “NoMinorHom” were considered as polymorphic and high-

quality, and were retained for further analysis. The genotyping results were exported as the pre-MAKEPED LINKAGE pedigree format(Lathrop et al. 1984).

The genotyping data were imported into Plink (version 1.9) to test pedigree information. A complete linkage agglomerative clustering procedure based on pairwise identity-by-state (IBS) distance were carried out. Multidimensional scaling analysis were performed on the generated IBS pairwise distances matrix. The clustering results were then represented with the first two clustering dimensions. The pedigree information of outliers was identified and checked by subsequent outlier detection diagnostics, where a Z score was assigned to measure the distance between the outliers with the rest of the samples. Outlier samples were discarded once they were detected with significantly larger distances compared with the normal level.

Linkage Map construction

Linkage map was constructed using Lep-MAP2 software. First, the Filtering module was executed to filtering low quality and un-informative markers. Markers with missing values larger than 12 (about 10%) or minimum allele frequency (MAF) less than 6 (about 5%) in each family were discarded. Only markers that have 2 or more informative families were kept. A segregation distortion test was also performed to compare the offspring genotype distribution and the expected Mendelian proportions. Markers with significant segregation distortion were filtered out (χ^2 test, $P < 0.005$). SNP Markers were assigned into linkage groups (LGs) using the SeparateChromosomes module. LGs were formed according to the threshold of logarithm of the odds (LOD) score limit of 35 and minimum LG size of 10. The singular markers were then added to the established LGs

using the JoinSingles module with an LOD score limit of 10 and a minimum difference of 3 between the best LG and the second best LG of each joined marker.

Marker order of each LGs were determined by allowing different recombination probabilities in both sex. Genotyping data of markers from the four families were analyzed simultaneously. For markers that were not mutually informative in all the families, their genetic distance was assessed by using haplotypes as partial inheritance vectors (Rastas et al. 2013). Two-round of marker ordering procedures were carried out for better performance (Figure 6). In the first round, ten iteration ordering-steps were conducted in each reference family, the marker order with the best likelihood across the four reference families were exported. To reduce the computing burden, a missing rate of 5% was set when deciding whether two markers were duplicates. As the number of markers may beyond the resolution of recombination. There were a lot of markers stack up in the same locus on the genetic map, which may lead to recombination rate deformities. Therefore, the stacked markers as well as the duplicated markers were clustered and filtered. The marker with the most informative meiosis of one cluster was selected as the representative marker and retained for a new round of marker ordering. After the second round of ordering, all the previously identified duplicates and stacked markers were added to the adjacent locus and included for genetic distance calculation with Kosambi mapping function by taking account both male and female meiosis. Sex-specific recombination rates were then calculated with the same marker order. MapChart (version 2.3) were used to graphically present the genetic linkage map.

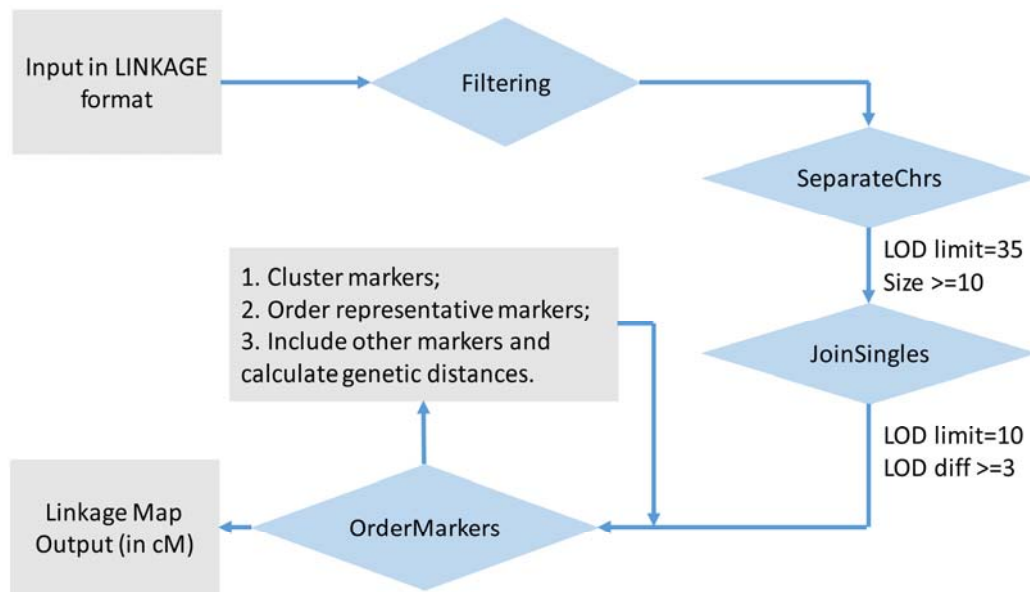


Figure 6. Genetic mapping procedure.

Results

Quality control for mapping families and SNP markers

SNP genotypes were obtained from 478 samples of four mapping families. By applying the criteria of DQC greater than 0.82 and call rate greater than 97%, 5 samples with pool qualities were recognized and filtered. Genotyping data of the rest 473 samples were imported into Plink for pedigree information test (Figure 7). The most majority of samples fall into three clusters, with two families gathered together because they were from one sire. The identified 8 outliers were also detected with abnormal ‘nearest neighbor’ scores, therefore, they were filtered from downstream analysis. The genotyping data of remaining 465 samples were imported into Lep-map2 for SNP filtering prior to linkage group assignment. According to the assessment of genotyping quality and polymorphism

in all samples from the four reference families, a total of 287,583 SNPs were informative in at least two families (Figure 8). For SNPs with identical genotypes in all mapping individuals, only one of which were kept for further analysis.

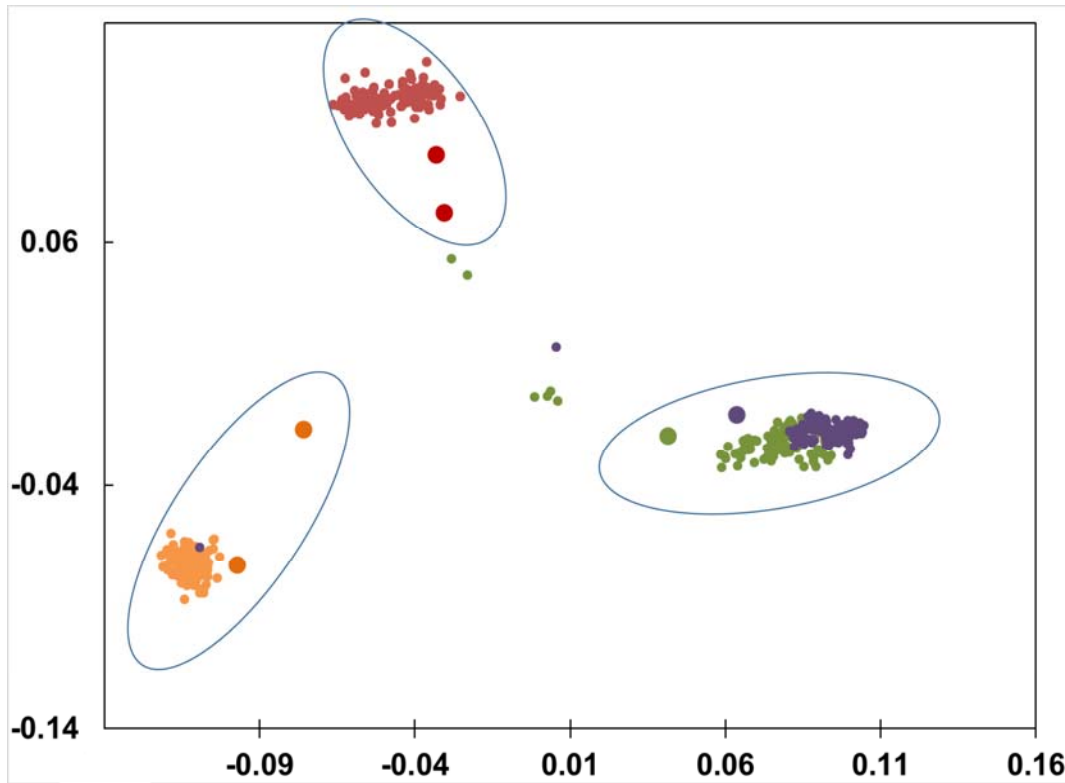


Fig 7. Sample structure identified by multidimensional scaling analysis of IBS distances. The coordinates were the first two clustering dimensions. Larger markers represent parent samples.

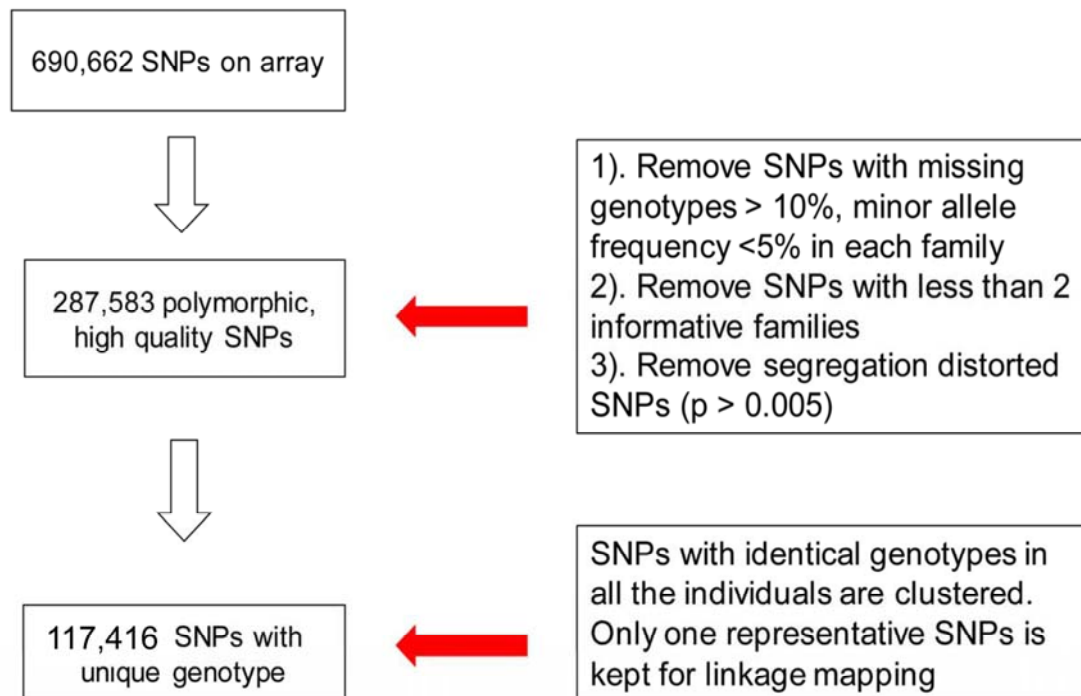


Figure 8. SNPs filtering process for linkage mapping

Linkage mapping

The qualified markers were first assigned into separate linkage groups with the SeparateChromosomes module. a total of 287,583 SNPs were assigned in to a total of 29 linkage groups, which was in concordance with the number of chromosomes of the catfish haploid genome. A two-round marker ordering procedure were carried out with the four families simultaneously. The first round of marker ordering identified 116,864 representative markers. By using the hidden Markov model (HMM), the OrderMarkers module is able to model recombinant haplotypes and identify markers with significant position difference across different mapping families. After filtering these unstable markers, a second round marker ordering step were performed to improve the order of representative markers. Finally, the previously excluded duplicate and stacked markers

were insert back into the maker order to calculate the genetic distance. Finally, a total of 253,744 markers were placed onto the new linkage map.

The sex-average genetic distances were calculated by taking account the recombination probabilities in both sexes. As summarized in Table 7, the sex-average map consists of 253,744 markers including 30,615 unique positions, with a total genetic length of 3,004.7 cM. The marker intervals estimated based on the unique marker positions ranged from 0.08 cM/marker pair in LG12 and LG13 to 0.13 cM/marker pair in LG22, with an average marker interval of 0.1 cM/marker pair in sex-average genetic map.

The markers on the sex-average map is well distributed, with no obvious gaps (Figure 9).

Table 7. Summary of the sex-average linkage map of channel catfish.

Sex-average map				
Linkage group	Mapped markers	Unique positions	Genetic length	Marker interval
1	13,473	1,198	122.48	0.1
2	6,639	1,031	97.624	0.09
3	12,864	1,342	136.613	0.1
4	7,769	1,090	111.502	0.1
5	9,882	1,017	91.712	0.09
6	11,557	1,314	126.625	0.1
7	8,013	832	83.681	0.1
8	11,036	1,218	112.518	0.09

9	8,538	1,269	122.109	0.1
10	6,143	850	80.721	0.09
11	9,282	1,364	135.765	0.1
12	8,293	1,011	84.4	0.08
13	9,362	1,287	109.138	0.08
14	6,744	899	86.76	0.1
15	7,704	1,000	92.755	0.09
16	9,309	1,181	126.838	0.11
17	10,062	1,287	114.924	0.09
18	13,500	1,228	113.228	0.09
19	7,449	1,040	89.178	0.09
20	10,846	1,050	110.144	0.1
21	10,109	1,021	91.477	0.09
22	5,637	773	103.989	0.13
23	10,508	996	98.423	0.1
24	5,378	778	88.577	0.11
25	7,894	996	101.438	0.1
26	7,173	864	91.43	0.11
27	7,935	864	86.565	0.1
28	6,099	1,068	104.627	0.1
29	4,546	747	89.494	0.12
Total	253,744	30,615	3,004.735	0.1

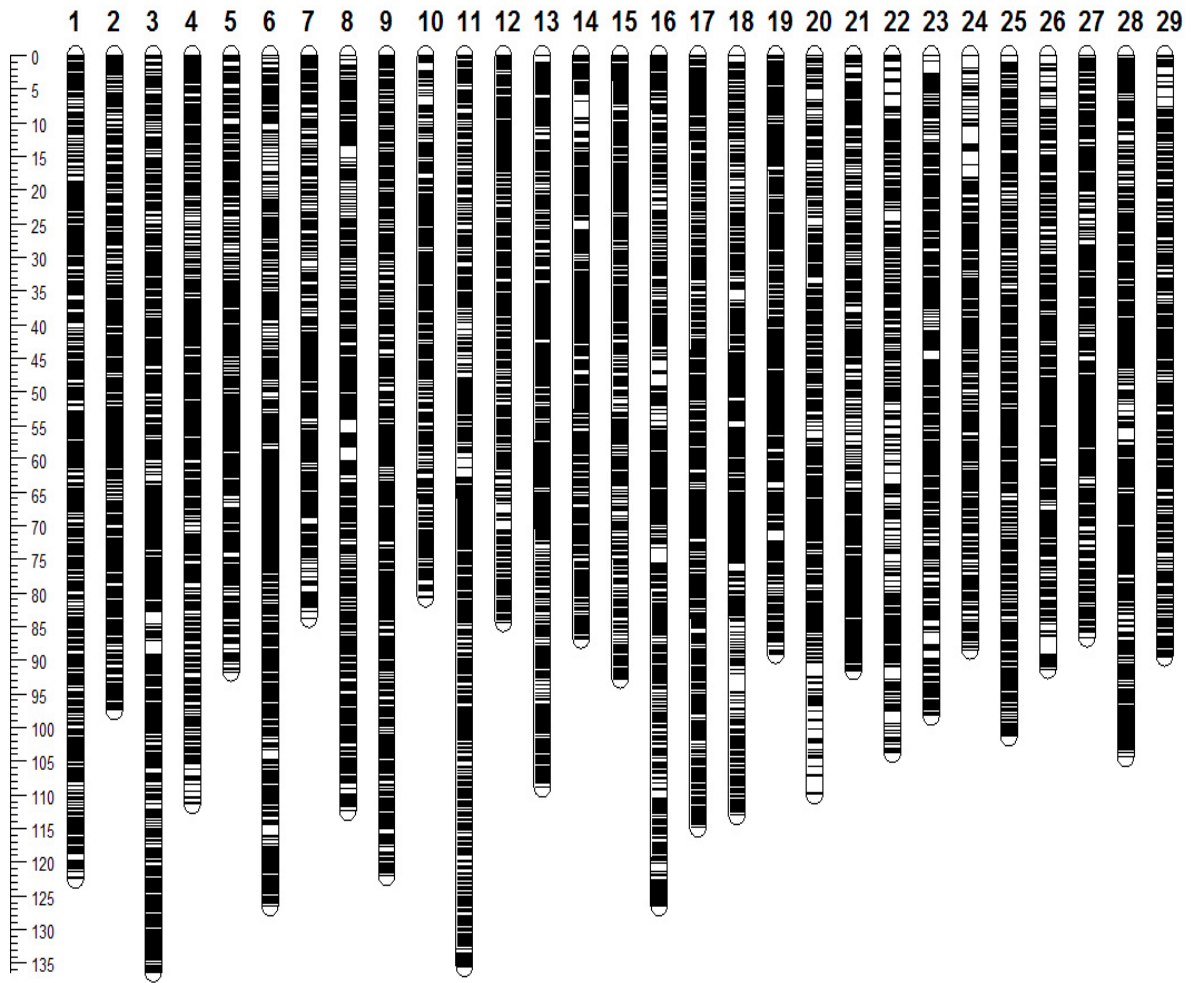


Figure 9. Illustration of sex-average linkage map

The sex-specific genetic distances were calculated by taking account the recombination probabilities in only one sex. The female genetic map consists of 23,610 unique markers, with a total genetic length of 3,582.26 cM (Table 8, Figure10). The marker intervals estimated based on the unique marker positions ranged from 0.13 cM/marker pair in LG5, LG13, LG19 and LG21 to 0.18 cM/marker in LG22 and LG29, with an average marker interval of 0.15 cM/marker in female genetic map. The male genetic map consists of 18,339 unique markers, with a total genetic length of 2,545.59 cM (Table 8, Figure11).

The marker intervals estimated based on the unique marker positions ranged from 0.11 cM/marker in LG12 to 0.17 cM/marker in LG22, with an average marker interval of 0.14 cM/marker in male genetic map.

The difference between female- and male-specific linkage map was assessed according to Ott's G-test (Ott, 1999). Significantly higher recombination rates were observed in a majority of the linkage groups of the female genetic map than that of male genetic map ($p < 0.01$). The female genetic map was 1,036.67 cM longer than male-genetic map, with an average female-to-male ratio of 1.4:1. The ratio varied by linkage groups, ranging from 0.96 in LG6 to 2.06 in LG18 (Table 8). Large differences of recombination rate were also observed in LG2 and LG23 with the recombination ratios of female to male greater than 1.6.

Table 8. Summary of the sex-specific linkage map of channel catfish.

Linkage group	Female-specific map			Male-specific map			F:M ratio
	Unique positions	Genetic length	Marker interval	Unique positions	Genetic length	Marker interval	
1	880	150.398	0.17	764	96.932	0.13	1.55
2	914	128.927	0.14	576	80.311	0.14	1.61
3	1,005	167.41	0.17	797	107.838	0.14	1.55
4	798	133.264	0.17	608	94.902	0.16	1.4
5	797	107.165	0.13	564	76.823	0.14	1.39
6	1,020	147.84	0.14	1,088	154.592	0.14	0.96
7	593	98.666	0.17	515	69.364	0.13	1.42

8	951	139.04	0.15	642	89.167	0.14	1.56
9	1,006	150.924	0.15	754	97.856	0.13	1.54
10	643	94.239	0.15	498	72.248	0.15	1.3
11	1,012	145.177	0.14	883	127.862	0.14	1.14
12	718	97.291	0.14	640	72.927	0.11	1.33
13	975	123.13	0.13	663	95.595	0.14	1.29
14	690	102.484	0.15	517	72.634	0.14	1.41
15	754	112.926	0.15	607	73.845	0.12	1.53
16	939	154.868	0.16	639	101.333	0.16	1.53
17	924	134.912	0.15	760	96.545	0.13	1.4
18	1,030	152.931	0.15	610	74.318	0.12	2.06
19	814	103.42	0.13	552	75.193	0.14	1.38
20	773	134.035	0.17	592	84.287	0.14	1.59
21	753	94.889	0.13	612	89.754	0.15	1.06
22	700	128.097	0.18	502	87.469	0.17	1.46
23	808	126.172	0.16	578	75.567	0.13	1.67
24	632	104.576	0.17	462	72.957	0.16	1.43
25	693	113.383	0.16	662	89.252	0.13	1.27
26	637	104.682	0.16	504	78.919	0.16	1.33
27	777	108.956	0.14	528	68.471	0.13	1.59
28	755	112.341	0.15	697	98.77	0.14	1.14
29	619	110.117	0.18	525	69.86	0.13	1.58
Total	23,610	3,582.26	0.15	18,339	2,545.591	0.14	1.41



Figure 10. Illustration of female-specific linkage map

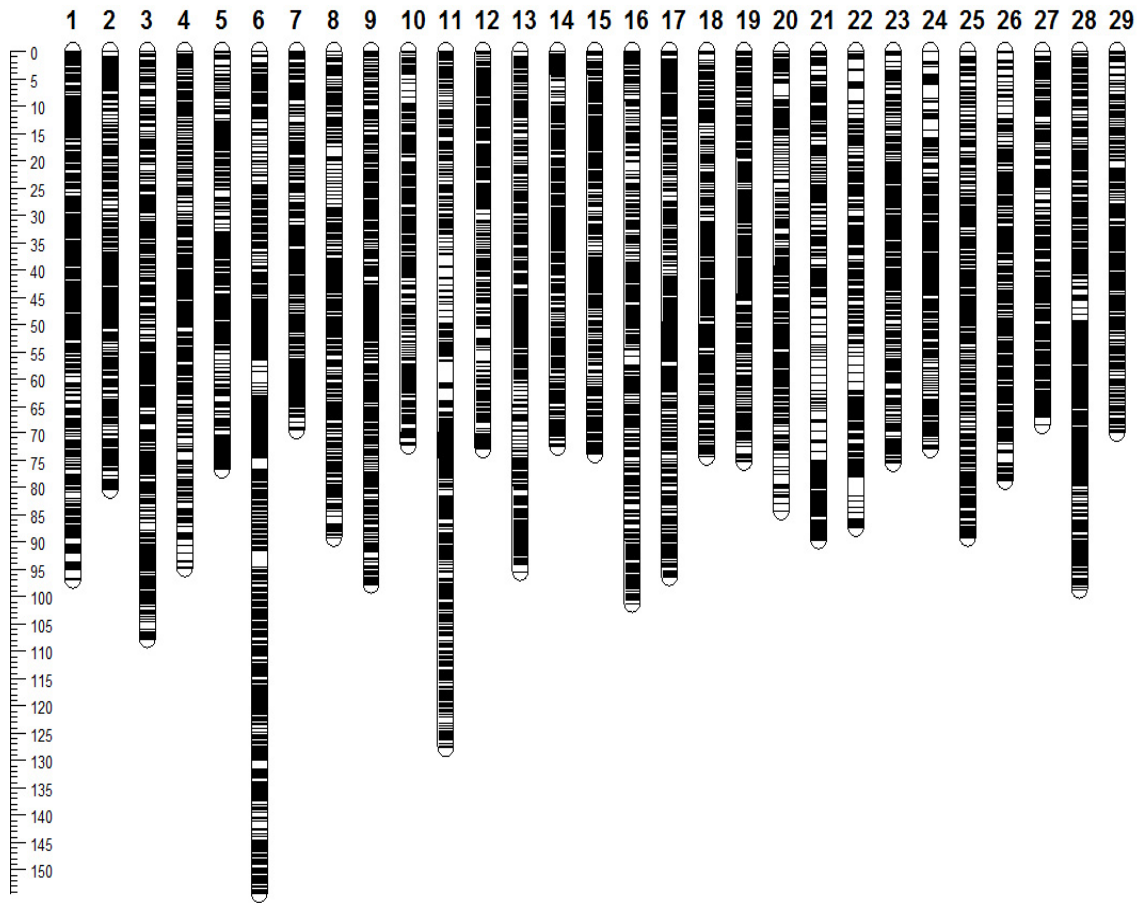


Figure 10. Illustration of male-specific linkage map

Integration and validation with whole genome sequence contigs

The genetic map anchored 1,546 out of 9,974 whole-genome sequencing scaffolds (Liu et al., 2016), corresponding to 766 Mb (97.8%) of the total 783 Mb. Compared with the recently generated reference genome sequence, a total of 5.7 Mb from 950 previously unmapped scaffolds were successful anchored to their corresponding positions (Figure 11). Additionally, 15 transcripts generated from de novo transcriptome assembly and 245 previously unmapped BAC-based physical contigs were also placed onto the linkage map. The positions of these contigs were illustrated in Figure 11.

As shown in Figure 11, mild to strong localized specific recombination pattern were observed in each linkage group. The recombination rates were usually elevated towards the ends and decreased in the middle of the chromosome. Stacked markers that located in regions of strong linkage disequilibrium were observed in each linkage group, especially in positions close to the centromeres rather than to the telomeres.

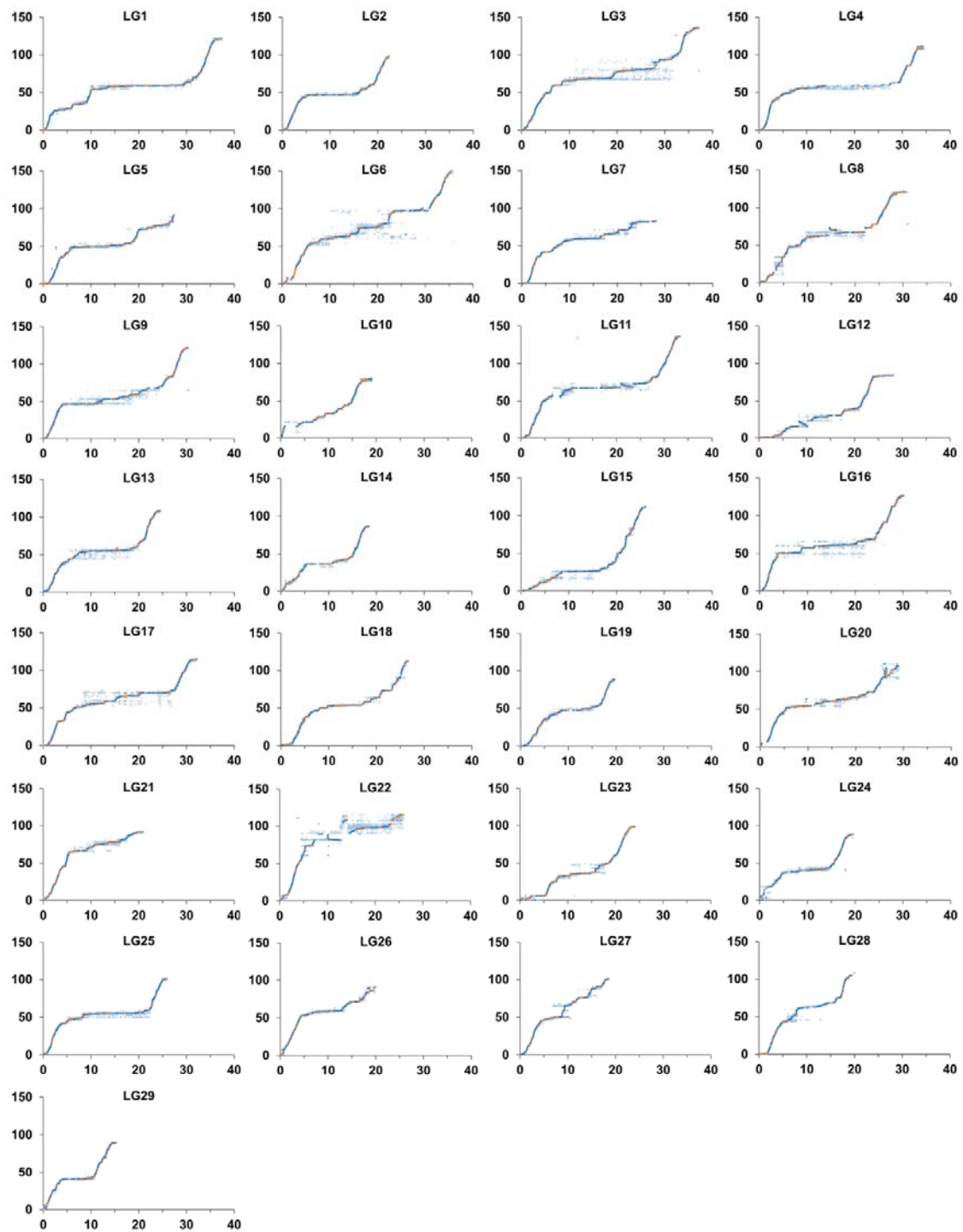


Figure 11. Concordance of SNP marker positions on reference sequence with those on genetic linkage map. The x-axis represents the markers physical position (Mb), the Y-

axis represents the marker's position on the linkage map (cM). Markers in blue were from previously mapped scaffolds, markers in orange were from previously unmapped contigs.

Discussion

In the present work, we constructed a high-density genetic map for channel catfish by using four reference families with a total of 466 individuals. This map possesses the highest marker density among all the genetic linkage maps constructed for aquaculture species. Taking advantage of the catfish 690K SNP array, we were able to efficiently genotype large number of samples for genetic linkage mapping. Such a high-density linkage map should be valuable resource for analysis and fine scale QTL mapping in catfish.

A large number of stacked markers were observed in the linkage map, reflecting the lack of recombination among these markers during meiosis. The only way to increase resolution power is to increase the sample sizes of the reference population (individuals of the reference families). While that can be done, the primary limitation is the cost. Each 690K array cost over \$150, and 1,000 samples would cost \$150,000, without considering all other cost including fish production, culture, sampling, DNA preparation, and genotyping. Due to the presence of genotyping error and missing values, genetic mapping with these closely linked markers greatly increase the computation burden and usually introduce an overestimated genetic sizes. In our study, the total genetic distance and error estimate dramatically reduced when we filtered the stacked SNPs. Genotyping errors in duplicate markers could also deform the marker orders, which can cause more severe problems compared with the inflation of genetic distances. To reduce the effect of clustered markers on the map construction and reduce the computing burden, we selected one representative marker to anchor the clusters. After the marker order is settled, the

duplicated markers were then rejoined into the linkage group. This procedure rescued the informative markers onto the linkage maps based on the positions of their representative anchor marker. Such clustered markers were still valuable to organize the scaffolds to chromosomes using the reference genome sequence.

By integration the genetic map and the reference genome sequence, the patterns of recombination across the whole chromosomes could be determined. Most often, the regions with strong linkage disequilibrium were distributed close the centromeres on the chromosomes, which was in concordance with the previously studies in catfish (Liu et al., 2003; Li et al., 2015), as well as other aquaculture species, such as tilapia (Kocher et al., 1998), medaka (Naruse et al., 2000), Atlantic salmon (Lorenz et al., 2010), rainbow trout (Sakamoto et al., 2000). Although the ultimate mechanism is still elusive, increasing evidence supported the notion that multiple functional sequence motifs were involved in the recombination regulation (Ross et al. 2015). With the available of channel catfish reference genome, the next step of our study is to identify the sequence features within the recombination “hot zones”.

Recombination rate in females was 1.4-fold higher than that in males, which was consistent with previously studies in channel catfish (Kucuktas et al., 2009; Li et al., 2015). This phenomenon was usually observed in species with dissimilar sex chromosomes, with larger recombination rate in homogametic sex than in heterogametic sex, such as mice (Lynn et al. 2005), zebrafish (Singer et al., 2002), Atlantic salmon (Lien et al. 2011), rainbow trout (Sakamoto et al. 2000), European seabass (Chistiakov et al. 2005), silver carp (Guo et al. 2013), and grass carp (Xu et al., 2010). Hypothesis toward the sexual dimorphism in recombination fraction is that sex chromosomes in homogametic sex is

equal in size, therefore, recombination is more likely to occur. However, this does not seem to be true for channel catfish. Interestingly, our results showed that recombination rate of LG4, which corresponds to the sex chromosome, is similar to that of other linkage groups, perhaps with the exception of the sex determination region. This suggested that other factors such as male-specific selection (Mank 2009) or chromatin differences may account for this difference (Makova and Hardison 2015).

The high-density genetic map can be a useful resource for genetic and genomic studies. By integrating it with the reference genome assembly, a total of 5.6 Mb from 950 previously unmapped contigs were successfully anchored to their corresponding positions. In addition, 15 transcripts generated from de novo transcriptome assembly and 245 previously unmapped BAC-based physical contigs were also placed onto the linkage map. This is a great improvement on integration of the linkage map and the reference genome sequence, which is useful for further genomic studies, QTL analysis, and whole genome-based selection.

REFERENCES

- Argue BJ, Liu Z, Dunham RA (2003) Dress-out and fillet yields of channel catfish, *Ictalurus punctatus*, blue catfish, *Ictalurus furcatus*, and their F₁, F₂ and backcross hybrids. *Aquaculture* 228:81-90.
- Aslam ML, Bastiaansen JWM, Elferink MG, Megens HJ, Crooijmans RPMA, Blomberg L, Fleischer RC, Van Tassell CP, Sonstegard TS, Schroeder SG, Groenen MAM, Long JA (2012) Whole genome SNP discovery and analysis of genetic diversity in Turkey (*Meleagris gallopavo*). *Bmc Genomics* 13.
- Baird NA, Etter PD, Atwood TS, Currey MC, Shiver AL, Lewis ZA, Selker EU, Cresko WA, Johnson EA (2008) Rapid SNP Discovery and Genetic Mapping Using Sequenced RAD Markers. *Plos One* 3.
- Baranski M, Gopikrishna G, Robinson NA, Katneni VK, Shekhar MS, Shanmugakarthik J, Jothivel S, Gopal C, Ravichandran P, Kent M, Arnyasi M, Ponniah AG (2014) The Development of a High Density Linkage Map for Black Tiger Shrimp (*Penaeus monodon*) Based on cSNPs. *Plos One* 9.
- Beldade P, Saenko SV, Pul N, Long AD (2009) A Gene-Based Linkage Map for *Bicyclus anynana* Butterflies Allows for a Comprehensive Analysis of Synteny with the Lepidopteran Reference Genome. *PLoS genetics* 5.
- Ben-Avi L, Durst R, Shpitzen S, Leitersdorf E, Meiner V (2004) A lipoprotein E genotyping: accurate, simple, high throughput method using ABI Prism((R)) SNaPshot (TM) multiplex system. *J Alzheimers Dis* 6:497-501.

- Bergfelder-Druing S, Grosse-Brinkhaus C, Lind B, Erbe M, Schellander K, Simianer H, Tholen E (2015) A Genome-Wide Association Study in Large White and Landrace Pig Populations for Number Piglets Born Alive. *Plos One* 10.
- Boussaha M, Guyomard R, Cabau C, Esquerre D, Quillet E (2012) Development and characterisation of an expressed sequence tags (EST)-derived single nucleotide polymorphisms (SNPs) resource in rainbow trout. *Bmc Genomics* 13.
- Burri R, Nater A, Kawakami T, Mugal CF, Olason PI, Smeds L, Suh A, Dutoit L, Bures S, Garamszegi LZ, Hogner S, Moreno J, Qvarnstrom A, Ruzic M, Saether SA, Saetre GP, Torok J, Ellegren H (2015) Linked selection and recombination rate variation drive the evolution of the genomic landscape of differentiation across the speciation continuum of *Ficedula* flycatchers. *Genome Res* 25:1656-1665.
- Cargill M, Altshuler D, Ireland J, Sklar P, Ardlie K, Patil N, Lane CR, Lim EP, Kalyanaraman N, Nemesh J, Ziaugra L, Friedland L, Rolfe A, Warrington J, Lipshutz R, Daley GQ, Lander ES (1999) Characterization of single-nucleotide polymorphisms in coding regions of human genes. *Nat Genet* 22:231-238.
- Castano-Sanchez C, Fuji K, Ozaki A, Hasegawa O, Sakamoto T, Morishima K, Nakayama I, Fujiwara A, Masaoka T, Okamoto H, Hayashida K, Tagami M, Kawai J, Hayashizaki Y, Okamoto N (2010) A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *Bmc Genomics* 11.
- Charlier C, Coppieters W, Rollin F, Desmecht D, Agerholm JS, Cambisano N, Carta E, Dardano S, Dive M, Fasquelle C, Frennet JC, Hanset R, Hubin X, Jorgensen C, Karim L, Kent M, Harvey K, Pearce BR, Simon P, Tama N, Nie H, Vandeputte S, Lien S, Longeri M, Fredholm M, Harvey RJ, Georges M (2008) Highly effective

- SNP-based association mapping and management of recessive defects in livestock. *Nat Genet* 40:449-454.
- Chen K, Rajewsky N (2006) Natural selection on human microRNA binding sites inferred from SNP data. *Nat Genet* 38:1452-1456.
- Chistiakov DA, Hellemans B, Haley CS, Law AS, Tsigenopoulos CS, Kotoulas G, Bertotto D, Libertini A, Volckaert FAM (2005) A microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. *Genetics* 170:1821-1826.
- Collins FS, Brooks LD, Chakravarti A (1998) A DNA polymorphism discovery resource for research on human genetic variation. *Genome Res* 8:1229-1231.
- Crispim AC, Kelly MJ, Guimaraes SEF, Silva FFE, Fortes MRS, Wenceslau RR, Moore S (2015) Multi-Trait GWAS and New Candidate Genes Annotation for Growth Curve Parameters in Brahman Cattle. *Plos One* 10.
- Cui J, Wang HD, Liu SK, Zhu LF, Qiu XM, Jiang ZQ, Wang XL, Liu ZJ (2014) SNP Discovery from Transcriptome of the Swimbladder of *Takifugu rubripes*. *Plos One* 9.
- Dalma-Weiszhausz DD, Warrington J, Tanimoto EY, Miyada G (2006) The affymetrix GeneChip (R) platform: An overview. *Method Enzymol* 410:3-28.
- Decker JE, Pires JC, Conant GC, Mckay SD, Heaton MP, Chen KF, Cooper A, Vilkki J, Seabury CM, Caetano AR, Johnson GS, Brenneman RA, Hanotte O, Eggert LS, Wiener P, Kim JJ, Kim KS, Sonstegard TS, Van Tassell CP, Neiberghs HL, McEwan JC, Brauning R, Coutinho LL, Babar ME, Wilson GA, McClure MC, Rolf MM, Kim J, Schnabel RD, Taylor JF (2009) Resolving the evolution of extant and extinct

- ruminants with high-throughput phylogenomics. *P Natl Acad Sci USA* 106:18644-18649.
- Didion JP, Yang HN, Sheppard K, Fu CP, McMillan L, de Villena FPM, Churchill GA (2012) Discovery of novel variants in genotyping arrays improves genotype retention and reduces ascertainment bias. *Bmc Genomics* 13.
- Djari A, Esquerre D, Weiss B, Martins F, Meersseman C, Boussaha M, Klopp C, Rocha D (2013) Gene-based single nucleotide polymorphism discovery in bovine muscle using next-generation transcriptomic sequencing. *Bmc Genomics* 14.
- Dunham R (2007) Comparison of six generations of selection, interspecific hybridization, intraspecific crossbreeding and gene transfer for growth improvement in ictalurid catfish. *Aquaculture* 272:S252-S253.
- Dunham R, Smitherman R (1982) Crossbreeding channel catfish for improvement of body weight in earthen ponds. *Growth* 47:97-103.
- Dunham RA, Argue BJ (1998) Seinability of channel catfish, blue catfish, and their F1, F2, F3, and backcross hybrids in earthen ponds. *The Progressive Fish-Culturist* 60:214-220.
- Dunham RA, Argue BJ (2000) Reproduction among channel catfish, blue catfish, and their F1 and F2 hybrids. *Transactions of the American Fisheries Society* 129:222-231.
- Dunham RA, Smitherman RO (1983) Response to selection and realized heritability for body weight in three strains of channel catfish, *Ictalurus punctatus*, grown in earthen ponds. *Aquaculture* 33:89-96.
- Dunham RA, Smitherman RO (1984) Ancestry and breeding of catfish in the United States.

- Dupree H, Green O (1969) Comparison of feed conversion and growth rate of six catfish species and their hybrids. Southeast Fish Cult Lab, Marion, AL.
- Dupree H, Green O, Sneed K (1969) Techniques for the Hybridization of Catfishes. US Fish and Wildlife Service, Southeastern Fish Cultural Laboratory, Publication 221.
- Dupuis MC, Zhang ZY, Druet T, Denoix JM, Charlier C, Lekeux P, Georges M (2011) Results of a haplotype-based GWAS for recurrent laryngeal neuropathy in the horse. *Mamm Genome* 22:613-620.
- Eck SH, Benet-Pages A, Flisikowski K, Meitinger T, Fries R, Strom TM (2009) Whole genome sequencing of a single *Bos taurus* animal for single nucleotide polymorphism discovery. *Genome Biology* 10.
- Eggen A (2012) The development and application of genomic selection as a new breeding paradigm. *Animal frontiers* 2:10-15.
- Elsik CG, Tellam RL, Worley KC, Gibbs RA, Abatepaulo ARR, Abbey CA, Adelson DL, Aerts J, Ahola V, Alexander L, Alioto T, Almeida IG, Amadio AF, Anatriello E, Antonarakis SE, Anzola JM, Astashyn A, Bahadue SM, Baldwin CL, Barris W, Baxter R, Bell SN, Bennett AK, Bennett GL, Biase FH, Boldt CR, Bradley DG, Brinkman FSL, Brinkmeyer-Langford CL, Brown WC, Brownstein MJ, Buhay C, Caetano AR, Camara F, Carroll JA, Carvalho WA, Casey T, Cervelatti EP, Chack J, Chacko E, Chandrabose MM, Chapin JE, Chapple CE, Chen HC, Chen L, Cheng Y, Cheng Z, Childers CP, Chitko-McKown CG, Chiu R, Choi JW, Chrast J, Colley AJ, Connelley T, Cree A, Curry S, Dalrymple B, Diep Dao M, Davis C, de Oliveira CJF, de Miranda Santos IKF, de Campos TA, Deobald H, Devinoy E, Dickens CM, Ding Y, Dinh HH, De Donato M, Donohue KE, Donthu R, Dovc P, Dugan-Rocha

S, Durbin KJ, Eberlein A, Edgar RC, Egan A, Eggen A, Eichler EE, Elhaik E, Ellis SA, Elnitski L, Ermolaeva O, Eyraas E, Fitzsimmons CJ, Fowler GR, Franzin AM, Fritz K, Gabisi RA, Garcia GR, Garcia JF, Genini S, Gerlach D, German JB, Gilbert JGR, Gill CA, Gladney CJ, Glass EJ, Goodell J, Grant JR, Graur D, Greaser ML, Green JA, Green RD, Guan L, Guigo R, Hadsell DL, Hagen DE, Hakimov HA, Halgren R, Hamernik DL, Hamilton C, Harhay GP, Harrow JL, Hart EA, Hastings N, Havlak P, Henrichsen CN, Hernandez J, Hernandez M, Herzig CTA, Hiendleder SG, Hines S, Hitchens ME, Hlavina W, Hobbs M, Holder M, Holt RA, Hu ZL, Hume J, Iivanainen A, Ingham A, Iso-Touru T, Jamis C, Jann O, Jensen K, Jhangiani SN, Jiang HY, Johnson AJ, Jones SJM, Joshi V, Junier T, Kapetis D, Kappes SM, Kapustin Y, Keele JW, Kent MP, Kerr T, Khalil SS, Khatib H, Kiryutin B, Kitts P, Kokocinski F, Kolbehdari D, Kovar CL, Kriventseva EV, Kumar CG, Kumar D, Lahmers KK, Landrum M, Larkin DM, Lau LPL, Leach R, Lee JCM, Lee S, Lemay DG, Lewin HA, Lewis LR, Li CX, Lien S, Liu GE, Liu YS, Liu Y, Logan KM, Lopez J, Lozado RJ, Lutzow YS, Lynn DJ, MacNeil MD, Maglott D, Malinverni R, Maqbool NJ, Marques E, Marra MA, Martin WF, Martins NF, Maruyama SR, Matukumalli LK, Mazza R, McEwan JC, McKay SD, McLean KL, McWilliam S, Medrano JF, Memili E, Moen C, Molenaar AL, Moore SS, Moore R, More DD, Moreno BT, Morgan MB, Muntean CT, Muzny DM, Nandakumar HP, Nazareth LV, Nguyen NB, Nicholas FW, Nogueira MFG, Okwuonu GO, Olsaker I, Pant SD, Panzitta F, Pastor RCP, Patel BM, Payne GM, Plass M, Poli MA, Poslusny N, Pruitt K, Pu LL, Qin X, Rachagani S, Raison JM, Ranganathan S, Ratnakumar A, Razpet A, Reecy J, Reese JT, Ren Y, Reymond A,

Riggs PK, Rijnkels M, Rincon G, Roberts A, Rodriguez-Osorio N, Rodriguez-Zas SL, Romero NE, Rosenwald A, Ruiz SJ, Sabo A, Salih H, Sando L, Santibanez J, Sapojnikov V, Schein JE, Schmutz SM, Schnabel RD, Schook L, Searle SM, Seo SW, Shen YF, Shen LB, Sherman L, Skow LC, Smith T, Snelling WM, Sodergren E, Solovyev V, Song H, Song JZ, Sonstegard TS, Southey BR, Souvorov A, Spurlock D, Steffen D, Stone RT, Stothard P, Sugimoto Y, Sweedler JV, Takasuga A, Tammen I, Taniguchi M, Taylor JF, Telugu BPVL, Ucla C, Urbanski JM, Utsunomiya YT, Van CP, Vattathil S, Verschoor CP, Waardenberg AJ, Walker A, Wang ZQ, Ward R, Warren JT, Waterman RC, Weikard R, Weinstock GM, Welsh TH, Wheeler DA, Wheeler TT, White SN, Whitside MD, Wilczek-Boney K, Williams JL, Williams RL, Wilming LG, Womack J, Wright RA, Wunderlich KR, Wyss C, Yang MQ, Yang JQ, Zdobnov EM, Zhang JK, Zhao FQ, Zhu B, Consortium BGSA (2009) The Genome Sequence of Taurine Cattle: A Window to Ruminant Biology and Evolution. *Science* 324:522-528.

Fahrenkrug SC, Freking BA, Smith TPL, Rohrer GA, Keele JW (2002) Single nucleotide polymorphism (SNP) discovery in porcine expressed genes. *Animal genetics* 33:186-195.

Frazer KA, Murray SS, Schork NJ, Topol EJ (2009) Human genetic variation and its contribution to complex traits. *Nat Rev Genet* 10:241-251.

Gabriel S, Ziaugra L, Tabbaa D (2009) SNP genotyping using the Sequenom MassARRAY iPLEX platform. *Current protocols in human genetics*:2.12. 11-12.12. 16.

- Gao ZX, Luo W, Liu H, Zeng C, Liu XL, Yi SJ, Wang WM (2012) Transcriptome Analysis and SSR/SNP Markers Information of the Blunt Snout Bream (*Megalobrama amblycephala*). *Plos One* 7.
- Geng X, Sha J, Liu SK, Bao LS, Zhang JR, Wang RJ, Yao J, Li C, Feng JB, Sun FY, Sun LY, Jiang C, Zhang Y, Chen AL, Dunham R, Zhi DG, Liu ZJ (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.
- Gibbs RA, Taylor JF, Van Tassell CP, Barendse W, Eversoe KA, Gill CA, Green RD, Hamernik DL, Kappes SM, Lien S, Matukumalli LK, McEwan JC, Nazareth LV, Schnabel RD, Taylor JF, Weinstock GM, Wheeler DA, Ajmone-Marsan P, Barendse W, Boettcher PJ, Caetano AR, Garcia JF, Hanotte O, Mariani P, Skow LC, Williams JL, Caetano AR, Diallo B, Green RD, Hailemariam L, Hanotte O, Martinez ML, Morris CA, Silva LOC, Spelman RJ, Taylor JF, Mulatu W, Zhao KY, Abbey CA, Agaba M, Araujo FR, Bunch RJ, Burton J, Gill CA, Gorni C, Olivier H, Harrison BE, Luff B, Machado MA, Mariani P, Morris CA, Mwakaya J, Plastow G, Sim W, Skow LC, Smith T, Sonstegard TS, Spelman RJ, Taylor JF, Thomas MB, Valentini A, Williams P, Womack J, Wooliams JA, Liu Y, Qin X, Worley KC, Gao C, Gill CA, Jiang HY, Liu Y, Moore SS, Nazareth LV, Ren YR, Song XZ, Bustamante CD, Hernandez RD, Muzny DM, Nazareth LV, Patil S, Lucas AS, Fu Q, Kent MP, Moore SS, Vega R, Abbey CA, Gao C, Gill CA, Green RD, Matukumalli LK, Matukumalli A, McWilliam S, Schnabel RD, Sclep G, Ajmone-Marsan P, Bryc K, Bustamante CD, Choi J, Gao H, Grefenstette JJ, Murdoch B, Stella A, Villa-Angulo R, Wright M, Aerts J, Jann O, Negrini R, Sonstegard TS,

- Williams JL, Taylor JF, Villa-Angulo R, Goddard ME, Hayes BJ, Barendse W, Bradley DG, Boettcher PJ, Bustamante CD, da Silva MB, Lau LPL, Liu GE, Lynn DJ, Panzitta F, Sclep G, Wright M, Dodds KG (2009) Genome-Wide Survey of SNP Variation Uncovers the Genetic Structure of Cattle Breeds. *Science* 324:528-532.
- Giudice JJ (1966) Growth of a blue x channel catfish hybrid as compared to its parent species. *The Progressive Fish-Culturist* 28:142-145.
- Goddard ME, Hayes BJ (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nat Rev Genet* 10:381-391.
- Goher M, Rauw W, Thin D, Gomez-Raya L (2010) Selective genotyping using genome-wide association studies (GWAS) that are associated with fiber diameter in Merino sheep. *J Dairy Sci* 93:169-169.
- Gonen S, Lowe NR, Cezard T, Gharbi K, Bishop SC, Houston RD (2014) Linkage maps of the Atlantic salmon (*Salmo salar*) genome derived from RAD sequencing. *Bmc Genomics* 15.
- Goudie CA, Simco BA, Davis KB (1995) Failure of gynogenetically-derived male channel catfish to produce all-male offspring. Paper presented at: Proceedings of the Fifth International Symposium on the Reproductive Physiology of Fish, Fish Symposium.
- Green O, Smitherman R, Pardue G (1976) Comparisons of growth and survival of channel catfish, *Ictalurus punctatus*, from distinct populations. Paper presented at: FAO, Rome (Italy) Fishery Resources and Environment Div FAO Technical Conference on Aquaculture Kyoto (Japan) 26 May 1976.

- Groenen MAM, Megens HJ, Zare Y, Warren WC, Hillier LW, Crooijmans RPMA, Vereijken A, Okimoto R, Muir WM, Cheng HH (2011) The development and characterization of a 60K SNP chip for chicken. *Bmc Genomics* 12.
- Guo WJ, Tong JG, Yu XM, Zhu CK, Feng X, Fu BD, He SP, Zeng FZ, Wang XH, Liu HY, Liu LS (2013) A second generation genetic linkage map for silver carp (*Hypophthalmichthys molitrix*) using microsatellite markers. *Aquaculture* 412:97-106.
- Gutierrez AP, Lubieniecki KP, Davidson EA, Lien S, Kent MP, Fukui S, Withler RE, Swift B, Davidson WS (2012) Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. *Aquaculture* 358:61-70.
- Guyomard R, Boussaha M, Krieg F, Hervet C, Quillet E (2012) A synthetic rainbow trout linkage map provides new insights into the salmonid whole genome duplication and the conservation of synteny among teleosts. *Bmc Genet* 13.
- Hanson T (2015) US Farm-Raised Catfish Industry 2014 Review and 2015 Outlook.
- Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819-1829.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. *J Dairy Sci* 92:433-443.
- He C, Chen L, Simmons M, Li P, Kim S, Liu ZJ (2003) Putative SNP discovery in interspecific hybrids of catfish by comparative EST analysis. *Animal genetics* 34:445-448.

- Hedgecock D, Shin G, Gracey AY, Van Den Berg D, Samanta MP (2015) Second-Generation Linkage Maps for the Pacific Oyster *Crassostrea gigas* Reveal Errors in Assembly of Genome Scaffolds. *G3-Genes Genom Genet* 5:2007-2019.
- Helyar SJ, Hemmer-Hansen J, Bekkevold D, Taylor MI, Ogden R, Limborg MT, Cariani A, Maes GE, Diopere E, Carvalho GR, Nielsen EE (2011) Application of SNPs for population genetics of nonmodel organisms: new opportunities and challenges. *Molecular ecology resources* 11:123-136.
- Helyar SJ, Limborg MT, Bekkevold D, Babbucci M, van Houdt J, Maes GE, Bargelloni L, Nielsen RO, Taylor MI, Ogden R, Cariani A, Carvalho GR, Panitz F, Consortium F (2012) SNP Discovery Using Next Generation Transcriptomic Sequencing in Atlantic Herring (*Clupea harengus*). *Plos One* 7.
- Holliday R, Grigg G (1993) DNA methylation and mutation. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis* 285:61-67.
- Huang XH, Feng Q, Qian Q, Zhao Q, Wang L, Wang AH, Guan JP, Fan DL, Weng QJ, Huang T, Dong GJ, Sang T, Han B (2009) High-throughput genotyping by whole-genome resequencing. *Genome Res* 19:1068-1076.
- Karlsson S, Moen T, Lien S, Glover KA, Hindar K (2011) Generic genetic differences between farmed and wild Atlantic salmon identified from a 7K SNP-chip. *Molecular ecology resources* 11:247-253.
- Kent M, Hayes B, Xiang Q, Berg P, Gibbs R, Lien S (2009) Development of 16.5 K SNP-Chip for Atlantic salmon. Paper presented at: Plant & Animal Genomes XVII Conference, San Diego, CA.

- Kerstens HHD, Crooijmans RPMA, Veenendaal A, Dibbits BW, Chin-A-Woeng TFC, den Dunnen JT, Groenen MAM (2009) Large scale single nucleotide polymorphism discovery in unsequenced genomes using second generation high throughput sequencing technology: applied to turkey. *Bmc Genomics* 10.
- Khatkar MS, Zenger KR, Hobbs M, Hawken RJ, Cavanagh JA, Barris W, McClintock AE, McClintock S, Thomson PC, Tier B, Nicholas FW, Raadsma HW (2007) A primary assembly of a bovine haplotype block map based on a 15,036-single-nucleotide polymorphism panel genotyped in holstein-friesian cattle. *Genetics* 176:763-772.
- Kijas JW, Townley D, Dalrymple BP, Heaton MP, Maddox JF, McGrath A, Wilson P, Ingersoll RG, McCulloch R, McWilliam S, Tang D, McEwan J, Cockett N, Oddy VH, Nicholas FW, Raadsma H, Conso ISG (2009) A Genome Wide Survey of SNP Variation Reveals the Genetic Structure of Sheep Breeds. *Plos One* 4.
- Kim ES, Shi X, Cobanoglu O, Weigel K, Berger PJ, Kirkpatrick BW (2009) Refined mapping of twinning-rate quantitative trait loci on bovine chromosome 5 and analysis of insulin-like growth factor-1 as a positional candidate gene. *J Anim Sci* 87:835-843.
- 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.
- Kumar S, You FM, Cloutier S (2012) Genome wide SNP discovery in flax through next generation sequencing of reduced representation libraries. *Bmc Genomics* 13.
- LaFramboise T (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Res* 37:4181-4193.

- Laghari MY, Lashari P, Zhang X, Xu P, Narejo NT, Xin B, Zhang Y, Sun X (2015) QTL mapping for economically important traits of common carp (*Cyprinus carpio* L.). *Journal of applied genetics* 56:65-75.
- Lai E (2001) Application of SNP technologies in medicine: Lessons learned and future challenges. *Genome Res* 11:927-929.
- Lapegue S, Harrang E, Heurtebise S, Flahauw E, Donnadiou C, Gayral P, Ballenghien M, Genestout L, Barbotte L, Mahla R, Haffray P, Klopp C (2014) Development of SNP-genotyping arrays in two shellfish species. *Molecular ecology resources* 14:820-830.
- Lathrop GM, Lalouel JM, Julier C, Ott J (1984) Strategies for Multilocus Linkage Analysis in Humans. *P Natl Acad Sci-Biol* 81:3443-3446.
- Leary JL (1908) Description of San Marcos Station with Some of the Methods of Propagation in use at That Station. *Transactions of the American Fisheries Society* 37:75-81.
- Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD (2005) A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170:237-244.
- Li C, Waldbieser G, Bosworth B, Beck BH, Thongda W, Peatman E (2014) SNP discovery in wild and domesticated populations of blue catfish, *Ictalurus furcatus*, using genotyping-by-sequencing and subsequent SNP validation. *Molecular ecology resources* 14:1261-1270.

- Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, Genome Project Data Processing S (2009) The Sequence Alignment/Map format and SAMtools. *Bioinformatics* 25:2078-2079.
- Li RW, Waldbieser GC (2006) Production and utilization of a high-density oligonucleotide microarray in channel catfish, *Ictalurus punctatus*. *Bmc Genomics* 7:134.
- Li Y, Liu S, Qin Z, Waldbieser G, Wang R, Sun L, Bao L, Danzmann RG, Dunham R, Liu Z (2015) Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. *DNA Research* 22:39-52.
- Lien S, Gidskehaug L, Moen T, Hayes BJ, Berg PR, Davidson WS, Omholt SW, Kent MP (2011) A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. *Bmc Genomics* 12.
- Lilyestrom C, Wolters W, Bury D, Rezk M, Dunham R (1999) Growth, carcass traits, and oxygen tolerance of diploid and triploid catfish hybrids. *North American journal of aquaculture* 61:293-303.
- Liu L, Li YH, Li SL, Hu N, He YM, Pong R, Lin DN, Lu LH, Law M (2012) Comparison of Next-Generation Sequencing Systems. *J Biomed Biotechnol*.
- Liu S, Li Y, Qin Z, Geng X, Bao L, Kaltenboeck L, Kucuktas H, Dunham R, Liu Z (2016a) High-density interspecific genetic linkage mapping provides insights into genomic incompatibility between channel catfish and blue catfish. *Animal genetics* 47:81-90.

- Liu S, Sun L, Li Y, Sun F, Jiang Y, Zhang Y, Zhang J, Feng J, Kaltenboeck L, Kucuktas H, Liu Z (2014) Development of the catfish 250K SNP array for genome-wide association studies. *BMC research notes* 7:135.
- Liu SK, Zhou ZC, Lu JG, Sun FY, Wang SL, Liu H, Jiang YL, Kucuktas H, Kaltenboeck L, Peatman E, Liu ZJ (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *Bmc Genomics* 12.
- Liu YL, Fann CSJ, Liu CM, Wu JY, Hung SL, Chan HY, Chen JJ, Pan CC, Liu SK, Hsieh MH, Hwang TJ, OuYang WC, Chen CY, Lin JJ, Chou FHC, Chueh CM, Liu WM, Tsuang MM, Faraone SV, Tsuang MT, Chen WJ, Hwu HG (2006) Absence of significant associations between four AKT1 SNP markers and schizophrenia in the Taiwanese population. *Psychiat Genet* 16:39-41.
- 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.
- Liu Z, Liu S, Yao J, Bao L, Zhang J, Li Y, Jiang C, Sun L, Wang R, Zhang Y, Zhou T, Zeng Q, et al. (2016b) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature communications* 7:11757.
- Liu Z, Nichols A, Li P, Dunham RA (1998) Inheritance and usefulness of AFLP markers in channel catfish (*Ictalurus punctatus*) blue catfish (*I-furcatus*), and their F1, F2, and backcross hybrids. *Mol Gen Genet* 258:260-268.
- Liu ZJ, Tan G, Li P, Dunham RA (1999) Transcribed dinucleotide microsatellites and their associated genes from channel catfish *Ictalurus punctatus*. *Biochem Bioph Res Co* 259:190-194.

- Lorenz S, Brenna-Hansen S, Moen T, Roseth A, Davidson WS, Omholt SW, Lien S (2010) BAC-based upgrading and physical integration of a genetic SNP map in Atlantic salmon. *Animal genetics* 41:48-54.
- Luan T, Woolliams JA, Lien S, Kent M, Svendsen M, Meuwissen TH (2009) The accuracy of Genomic Selection in Norwegian red cattle assessed by cross-validation. *Genetics* 183:1119-1126.
- Lynn A, Schrupp S, Cherry J, Hassold T, Hunt P (2005) Sex, not genotype, determines recombination levels in mice. *The American Journal of Human Genetics* 77:670-675.
- Makova KD, Hardison RC (2015) The effects of chromatin organization on variation in mutation rates in the genome. *Nat Rev Genet* 16:213-223.
- Mank JE (2009) The evolution of heterochiasmy: the role of sexual selection and sperm competition in determining sex-specific recombination rates in eutherian mammals. *Genet Res* 91:355-363.
- Mardis ER (2008) The impact of next-generation sequencing technology on genetics. *Trends Genet* 24:133-141.
- Marklund S, Carlborg O (2010) SNP detection and prediction of variability between chicken lines using genome resequencing of DNA pools. *Bmc Genomics* 11.
- Matukumalli LK, Lawley CT, Schnabel RD, Taylor JF, Allan MF, Heaton MP, O'Connell J, Moore SS, Smith TPL, Sonstegard TS, Van Tassell CP (2009) Development and Characterization of a High Density SNP Genotyping Assay for Cattle. *Plos One* 4.
- McCue ME, Bannasch DL, Petersen JL, Gurr J, Bailey E, Binns MM, Distl O, Guerin G, Hasegawa T, Hill EW, Leeb T, Lindgren G, Penedo MC, Roed KH, Ryder OA,

- Swinburne JE, Tozaki T, Valberg SJ, Vaudin M, Lindblad-Toh K, Wade CM, Mickelson JR (2012) A high density SNP array for the domestic horse and extant *Perissodactyla*: utility for association mapping, genetic diversity, and phylogeny studies. *PLoS genetics* 8:e1002451.
- Miller JM, Poissant J, Kijas JW, Coltman DW, Consortium ISG (2011) A genome-wide set of SNPs detects population substructure and long range linkage disequilibrium in wild sheep. *Molecular ecology resources* 11:314-322.
- Miller S (2010) Genetic improvement of beef cattle through opportunities in genomics. *Revista Brasileira de Zootecnia* 39:247-255.
- Morin PA, Luikart G, Wayne RK, Grp SW (2004) SNPs in ecology, evolution and conservation. *Trends Ecol Evol* 19:208-216.
- Mulcahy TM, Kunkel MA, Greeny D, Liberchuk R, Camisa A, Studebaker J, Gelfand CA, Bell PA, Boyce-Jacino M (2001) SNPstream (R) MT, a microsphere-based, primer extension system for the analysis of single nucleotide polymorphisms. *Am J Hum Genet* 69:460-460.
- 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.
- Ohara E, Nishimura T, Nagakura Y, Sakamoto T, Mushiake K, Okamoto N (2005) Genetic linkage maps of two yellowtails (*Seriola quinqueradiata* and *Seriola lalandi*). *Aquaculture* 244:41-48.

- Oliphant A, Barker DL, Stuelpnagel JR, Chee MS (2002) BeadArray (TM) technology: Enabling an accurate, cost-effective approach to high throughput genotyping. *Biotechniques*:56-+.
- Palaiokostas C, Bekaert M, Davie A, Cowan ME, Oral M, Taggart JB, Gharbi K, McAndrew BJ, Penman DJ, Migaud H (2013) Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *Bmc Genomics* 14.
- Palti Y, Gao G, Liu S, Kent MP, Lien S, Miller MR, Rexroad CE, Moen T (2015) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular ecology resources* 15:662-672.
- Palti Y, Gao GT, Miller MR, Vallejo RL, Wheeler PA, Quillet E, Yao JB, Thorgaard GH, Salem M, Rexroad CE (2014) A resource of single-nucleotide polymorphisms for rainbow trout generated by restriction-site associated DNA sequencing of doubled haploids. *Molecular ecology resources* 14:588-596.
- Pant SD, Schenkel FS, Verschoor CP, You QM, Kelton DF, Moore SS, Karrow NA (2010) A principal component regression based genome wide analysis approach reveals the presence of a novel QTL on BTA7 for MAP resistance in holstein cattle. *Genomics* 95:176-182.
- Picoult-Newberg L, Ideker TE, Pohl MG, Taylor SL, Donaldson MA, Nickerson DA, Boyce-Jacino M (1999) Mining SNPs from EST databases. *Genome Res* 9:167-174.
- Ragoussis J (2009) Genotyping Technologies for Genetic Research. *Annu Rev Genom Hum G* 10:117-133.

- Ramos AM, Crooijmans RPMA, Affara NA, Amaral AJ, Archibald AL, Beever JE, Bendixen C, Churcher C, Clark R, Dehais P, Hansen MS, Hedegaard J, Hu ZL, Kerstens HH, Law AS, Megens HJ, Milan D, Nonneman DJ, Rohrer GA, Rothschild MF, Smith TPL, Schnabel RD, Van Tassell CP, Taylor JF, Wiedmann RT, Schook LB, Groenen MAM (2009) Design of a High Density SNP Genotyping Assay in the Pig Using SNPs Identified and Characterized by Next Generation Sequencing Technology. *Plos One* 4.
- Rastas P, Paulin L, Hanski I, Lehtonen R, Auvinen P (2013) Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* 29:3128-3134.
- Rexroad CE, Palti Y, Gahr SA, Vallejo RL (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *Bmc Genet* 9.
- Rhode C, Hepple JA, Jansen S, Davis T, Vervalle J, Bester-van der Merwe AE, Roodt-Wilding R (2012) A population genetic analysis of abalone domestication events in South Africa: Implications for the management of the abalone resource. *Aquaculture* 356:235-242.
- Rincon G, Weber KL, Van Eenennaam AL, Golden BL, Medrano JF (2011) Hot topic: Performance of bovine high-density genotyping platforms in Holsteins and Jerseys. *J Dairy Sci* 94:6116-6121.
- Roselius K, Stephan W, Stadler T (2005) The relationship of nucleotide polymorphism, recombination rate and selection in wild tomato species. *Genetics* 171:753-763.
- Ross CR, DeFelice DS, Hunt GJ, Ihle KE, Amdam GV, Rueppell O (2015) Genomic correlates of recombination rate and its variability across eight recombination maps in the western honey bee (*Apis mellifera* L.). *Bmc Genomics* 16.

- Sakamoto T, Danzmann RG, Gharbi K, Howard P, Ozaki A, Khoo SK, Woram RA, Okamoto N, Ferguson MM, Holm LE, Guyomard R, Hoyheim B (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* 155:1331-1345.
- Sattarzadeh A, Achenbach U, Lubeck J, Strahwald J, Tacke E, Hofferbert HR, Rothsteyn T, Gebhardt C (2006) Single nucleotide polymorphism (SNP) genotyping as basis for developing a PCR-based marker highly diagnostic for potato varieties with high resistance to *Globodera pallida* pathotype Pa2/3. *Mol Breeding* 18:301-312.
- Schmit SL, Gollub J, Shapero MH, Huang SC, Rennert HS, Finn A, Rennert G, Gruber SB (2015) MicroRNA Polymorphisms and Risk of Colorectal Cancer. *Cancer Epidem Biomar* 24:65-72.
- Shastry BS (2002) SNP alleles in human disease and evolution. *J Hum Genet* 47:561-566.
- Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovskiy R, Zilberman N, Kocher TD, Hulata G, Ron M (2006) *Amh* and *Dmrta2* genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* 174:1573-1581.
- Smitherman RO, Dunham R (1993) Relationships among cultured and naturally occurring populations of freshwater catfish in the United States. *Interactions Between Cultured Species and Naturally Occurring Species in the Environment*, Homer, Alaska, Elmer E Rasmuson Library Cataloging-In-Publication.
- Snelling WM, Allan MF, Keele JW, Kuehn LA, McDanel T, Smith TPL, Sonstegard TS, Thallman RM, Bennett GL (2010) Genome-wide association study of growth in crossbred beef cattle. *J Anim Sci* 88:837-848.

- Song WT, Li YZ, Zhao YW, Liu Y, Niu YZ, Pang RY, Miao GD, Liao XL, Shao CW, Gao FT, Chen SL (2012) Construction of a High-Density Microsatellite Genetic Linkage Map and Mapping of Sexual and Growth-Related Traits in Half-Smooth Tongue Sole (*Cynoglossus semilaevis*). *Plos One* 7.
- Stothard P, Choi JW, Basu U, Sumner-Thomson JM, Meng Y, Liao XP, Moore SS (2011) Whole genome resequencing of Black Angus and Holstein cattle for SNP and CNV discovery. *Bmc Genomics* 12.
- Sun LY, Liu SK, Wang RJ, Jiang YL, Zhang Y, Zhang JR, Bao LS, Kaltenboeck L, Dunham R, Waldbieser G, Liu ZJ (2014) Identification and Analysis of Genome-Wide SNPs Provide Insight into Signatures of Selection and Domestication in Channel Catfish (*Ictalurus punctatus*). *Plos One* 9.
- Swingle HS (1958) Experiments on growing fingerling channel catfish to marketable size in ponds.
- Taillon-Miller P, Gu ZJ, Li Q, Hillier L, Kwok PY (1998) Overlapping genomic sequences: A treasure trove of single-nucleotide polymorphisms. *Genome Res* 8:748-754.
- Tsigenopoulos CS, Louro B, Chatziplis D, Lagnel J, Vogiatzi E, Loukovitis D, La Franch R, Sarropoulou E, Power DM, Patarnello T, Mylonas CC, Magoulas A, Bargelloni L, Canario A, Kotoulas G (2014) Second generation genetic linkage map for the gilthead sea bream *Sparus aurata* L. *Mar Genom* 18:77-82.
- Van Tassell CP, Smith TPL, Matukumalli LK, Taylor JF, Schnabel RD, Lawley CT, Haudenschild CD, Moore SS, Warren WC, Sonstegard TS (2008) SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. *Nat Methods* 5:247-252.

- VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, Taylor JF, Schenkel FS (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci* 92:16-24.
- Wade CM, Giulotto E, Sigurdsson S, Zoli M, Gnerre S, Imsland F, Lear TL, Adelson DL, Bailey E, Bellone RR, Blocker H, Distl O, Edgar RC, Garber M, Leeb T, Mauceli E, MacLeod JN, Penedo MCT, Raison JM, Sharpe T, Vogel J, Andersson L, Antczak DF, Biagi T, Binns MM, Chowdhary BP, Coleman SJ, Della Valle G, Fryc S, Guerin G, Hasegawa T, Hill EW, Jurka J, Kiialainen A, Lindgren G, Liu J, Magnani E, Mickelson JR, Murray J, Nergadze SG, Onofrio R, Pedroni S, Piras MF, Raudsepp T, Rocchi M, Roed KH, Ryder OA, Searle S, Skow L, Swinburne JE, Syvanen AC, Tozaki T, Valberg SJ, Vaudin M, White JR, Zody MC, Lander ES, Lindblad-Toh K, Perfo BIGS, T BIWGA (2009) Genome Sequence, Comparative Analysis, and Population Genetics of the Domestic Horse. *Science* 326:865-867.
- Waldbieser GC, Bosworth BG (1997) Cloning and characterization of microsatellite loci in channel catfish, *Ictalurus punctatus*. *Animal genetics* 28:295-298.
- Waldbieser GC, Bosworth BG, Nonneman DJ, Wolters WR (2001) A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genetics* 158:727-734.
- Wang CM, Bai ZY, He XP, Lin G, Xia JH, Sun F, Lo LC, Feng F, Zhu ZY, Yue GH (2011) A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *Bmc Genomics* 12:174.
- Wang DG, Fan JB, Siao CJ, Berno A, Young P, Sapolsky R, Ghandour G, Perkins N, Winchester E, Spencer J, Kruglyak L, Stein L, Hsie L, Topaloglou T, Hubbell E,

- Robinson E, Mittmann M, Morris MS, Shen NP, Kilburn D, Rioux J, Nusbaum C, Rozen S, Hudson TJ, Lipshutz R, Chee M, Lander ES (1998) Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. *Science* 280:1077-1082.
- Wang L, Wan ZY, Bai B, Huang SQ, Chua E, Lee M, Pang HY, Wen YF, Liu P, Liu F, Sun F, Lin G, Ye BQ, Yue GH (2015a) Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Scientific reports* 5:16358.
- Wang S, Peatman E, Abernathy J, Waldbieser G, Lindquist E, Richardson P, Lucas S, Wang M, Li P, Thimmapuram J (2010) Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome biology* 11:1-14.
- Wang SL, Sha ZX, Sonstegard TS, Liu H, Xu P, Somridhivej B, Peatman E, Kucuktas H, Liu ZJ (2008) Quality assessment parameters for EST-derived SNPs from catfish. *Bmc Genomics* 9.
- Wang WH, Zhang T, Wang JY, Zhang GX, Wang YJ, Zhang YW, Zhang JH, Li GH, Xue Q, Han KP, Zhao XH, Zheng HK (2016) Genome-wide association study of 8 carcass traits in Jinghai Yellow chickens using specific-locus amplified fragment sequencing technology. *Poultry Sci* 95:500-506.
- Wang Y, Yang L, Wu B, Song Z, He S (2015b) Transcriptome analysis of the plateau fish (*Triplophysa dalaica*): Implications for adaptation to hypoxia in fishes. *Gene* 565:211-220.

- Weber CC, Boussau B, Romiguier J, Jarvis ED, Ellegren H (2014) Evidence for GC-biased gene conversion as a driver of between-lineage differences in avian base composition. *Genome Biology* 15.
- Wiedmann RT, Smith TPL, Nonneman DJ (2008) SNP discovery in swine by reduced representation and high throughput pyrosequencing. *Bmc Genet* 9.
- Wiggans GR, Sonstegard TS, Vanraden PM, Matukumalli LK, Schnabel RD, Taylor JF, Schenkel FS, Van Tassell CP (2009) Selection of single-nucleotide polymorphisms and quality of genotypes used in genomic evaluation of dairy cattle in the United States and Canada. *J Dairy Sci* 92:3431-3436.
- Wringe BF, Devlin RH, Ferguson MM, Moghadam HK, Sakhrani D, Danzmann RG (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *Bmc Genet* 11.
- Xu J, Ji PF, Zhao ZX, Zhang Y, Feng JX, Wang J, Li JT, Zhang XF, Zhao L, Liu GZ, Xu P, Sun XW (2012) Genome-Wide SNP Discovery from Transcriptome of Four Common Carp Strains. *Plos One* 7.
- Xu J, Zhao Z, Zhang X, Zheng X, Li J, Jiang Y, Kuang Y, Zhang Y, Feng J, Li C, Yu J, Li Q, Zhu Y, Liu Y, Xu P, Sun X (2014) Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *Bmc Genomics* 15:307.
- Yanez J (2014) Development of a 200K SNP array for Atlantic Salmon: exploiting across continents genetic variation. Paper presented at: 10th World Congress on Genetics Applied to Livestock Production (Asas).
- Yanez JM, Naswa S, Lopez ME, Bassini L, Correa K, Gilbey J, Bernatchez L, Norris A, Neira R, Lhorente JP, Schnable PS, Newman S, Mileham A, Deeb N, Di Genova

- A, Maass A (2016) Genome-wide single nucleotide polymorphism (SNP) discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular ecology resources*.
- You FM, Huo NX, Deal KR, Gu YQ, Luo MC, McGuire PE, Dvorak J, Anderson OD (2011) Annotation-based genome-wide SNP discovery in the large and complex *Aegilops tauschii* genome using next-generation sequencing without a reference genome sequence. *Bmc Genomics* 12.
- Yu H, You XX, Li J, Liu HK, Meng ZN, Xiao L, Zhang HF, Lin HR, Zhang Y, Shi Q (2016) Genome-Wide Mapping of Growth-Related Quantitative Trait Loci in Orange-Spotted Grouper (*Epinephelus coioides*) Using Double Digest Restriction-Site Associated DNA Sequencing (ddRADseq). *Int J Mol Sci* 17.
- Yue GH (2014) Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish Fish* 15:376-396.
- Zhou LC, Holliday JA (2012) Targeted enrichment of the black cottonwood (*Populus trichocarpa*) gene space using sequence capture. *Bmc Genomics* 13.
- Zhu C, Tong J, Yu X, Guo W, Wang X, Liu H, Feng X, Sun Y, Liu L, Fu B (2014) A second-generation genetic linkage map for bighead carp (*Aristichthys nobilis*) based on microsatellite markers. *Animal genetics* 45:699-708.