

Towards the identification of sex-determining gene(s) in channel catfish

by

Fanyue Sun

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

Auburn, Alabama
August 3rd, 2013

Keywords: sex determination, sex chromosome, next-generation sequencing,
genome assembly and scaffolding, comparative analysis, catfish

Copyright 2013 by Fanyue Sun

Approved by

Zhanjiang Liu, Chair, Alumni Professor of Fisheries and Allied Aquacultures
Joanna Wysocka-Diller, Associate Professor of Biological Sciences
Nannan Liu, Professor of Plant Pathology and Entomology
Rex Dunham, Alumni Professor of Fisheries and Allied Aquacultures

Abstract

Sex is the most fundamental feature in the life of an organism. Studying of sex determination is an important area in animal developmental and evolutionary biology, as well as ecology. In teleost, sex determination mechanism exhibits extraordinary plasticity and diversity with respect to the evolutionary pattern.

Catfish have a XY male/XX female sex chromosome system. The exact mechanism of sex determination in catfish is unknown at present. As a first step towards the identification of sex-determining genes in catfish, we performed the first transcriptome-level analysis of the catfish testis using high throughput Illumina sequencing to understand the transcriptome of the catfish testes. Gene ontology and annotation analysis suggested that many of the male-biased genes identified from the analysis were involved in gonadogenesis, spermatogenesis, testicular determination, gametogenesis, gonad differentiation, and possibly sex determination. Our analysis would lay the basis for further follow-up analysis of genes involved in sex determination and differentiation in catfish.

To move toward the goal of identification of the Y-specific fragments in channel catfish, we utilized multiple approaches to get the best assembly and scaffolding of X and Y chromosome in catfish and conducted *in silico* comparative analysis between these two chromosomes. Sequencing of the super-male (YY) sample, the regular male (XY) sample, and BAC clones on the sex chromosome using the Illumina HiSeq 2000 platform followed by assembly by ABySS and scaffolding by SSPACE. Comparative analyses were performed with

the rapidly genome aligning software MUMmer. However, by comparing the nucleotide sequences of the X-chromosome and Y-chromosome in detail, no Y-specific sequence was identified. The present work demonstrates that there is no gene deletion/insertion between X and Y chromosome in channel catfish.

The position of the putative catfish sex-determining region was defined by assuming that the previously identified SD marker is right in the middle of the two adjacent informative sex-linked markers in both directions with the lowest recombination frequencies. The BAC markers covering this sex-determining region were used as query to BLASTN-searched in the Y-chromosome scaffolds in order to obtain the full coverage of this region. Genes in this SD region were identified using BLASTX-searched in the non-redundant (nr) protein database for annotation. Thus, the SD region is 203 kbp in length, harboring 10 annotated genes. Gene expression in male vs. female embryos at the early gonad differentiation stages will be tested among these candidate genes.

Acknowledgments

I would love to express my deepest gratitude to my major professor, Dr. Zhanjiang (John) Liu, for his patience, support, and enlightening guidance during my Ph. D study. Thanks are also given to all of my committee members Dr. Rex Dunham, Dr. Nannan Liu, Dr. Joanna Wysocka-Diller and the outside reader Dr. Bernhard Kaltenboeck for their time and profound and carefully aimed comments on my dissertation. My sincere appreciation is extended to Dr. Eric Peatman, Dr. Huseyin Kucuktas, Ludmilla Kaltenboeck, and every member in our Fish Molecular Genetics and Biotechnology Laboratory and all the faculty and staff in the Department of Fisheries and Allied Aquacultures. Thank you very much for bringing me so many great memories. Above all, I would like to give special thanks to my gorgeous parents, for the continuous love, understanding and encouragement that have never been changed.

Table of Contents

Abstract	ii
Acknowledgments.....	iv
List of Tables	vi
List of Figures	vii
Chapter I. Literature review	1
Chapter II. Overview and research objectives	25
Chapter III. Male-biased genes in catfish as revealed by RNA-Seq analysis of the testis transcriptome	28
Chapter IV. Assembly and comparative analysis of the X and Y chromosomes in channel catfish, and identification of the putative catfish sex-linked region.....	55
Chapter V. Summary and future directions	78
Cumulative bibliography	82

List of Tables

Table 1-1 Previously reported candidate “master” sex-determining genes in vertebrates	7
Table 1-2 Comparison of several important aspects of the sequencing statistics in various mainstream platforms.....	21
Table 3-1 Summary of RNA-Seq of the testis transcriptome	34
Table 3-2 Assembly statistics of the testes transcriptome using Trans-ABYSS	35
Table 3-3 Representative channel catfish male-biased genes involved in spermatogenesis, gonad sex determination, and testicular determination.....	39
Table 4-1 Summary of Illumina sequencing data and metrics of the ABYSS <i>de novo</i> assembly results of male catfish resources.....	71
Table 4-2 Summary of the metrics of the male genome scaffolding results from five scaffolding schemes	72
Table 4-3 Summary of X chromosome and Y chromosome of channel catfish.....	74
Table 4-4 Gene list in the putative sex-linked region of channel catfish.....	77

List of Figures

Figure 3-1 GO term classifications of channel catfish testis transcriptome	37
Figure 3-2 Putative catfish spermatogenic pathway based on RNA-Seq expression signatures in channel catfish testes	41
Figure 3-3 Linkage group distribution of the male-biased genes	42
Figure 4-1 A schematic overview of catfish Y chromosome assembly and scaffolding.....	67
Figure 4-2 Output of MUMmer sequence alignment package to identify Y-specific regions in the large sequence sets on the Y chromosome	73
Figure 4-3 Partial linkage map of catfish LG 4 with the putative sex-linked region	76

CHAPTER I. LITERATURE REVIEW
SEX DETERMINATION IN FISH AND ADVANCED NEXT-GENERATION
SEQUENCING APPROACHES

Fish, aquaculture and channel catfish

According to the latest record in FishBase, the largest and most extensively accessed online database of fish species (specifically finfish), it is believed that fish represent over 50% of all vertebrates with over 32,500 species (FishBase, 2013). Fishes are the most evolutionarily diverse group of vertebrates by far in terms of their morphology, ecology, development, behavior and genomes and multiple other aspects of biology (Nelson 1994).

Fish play an important role in providing an alternative source of food. It is one of the most valuable agricultural contributions to the global economic growth, as well as to reducing poverty, increasing the growth and ensuring food and nutrition security. Aquaculture, which is probably the fastest growing food-producing sector, now makes up approximately half of the world's food fish (FAO, 2013). Thus, fish producers, like those of more traditional livestock, desire maximized improvement in fish production.

Channel cattish, *Ictalurus punctatus*, is the most important aquatic animal species in the United States. It belongs to the family Ictaluridae, order Siluriformes. Originally, channel catfish were found only in the Gulf States and the Mississippi Valley north to the prairie provinces of Canada and Mexico. Currently, it has been widely introduced throughout the United States and the world.

Channel catfish could live in various habitats, such as large reservoirs, lakes, ponds and some sluggish streams. They are usually found where bottoms are sand, gravel or rubble, in preference to mud bottoms. Generally, they prefer clear water streams. However, it is common to

culture catfish in muddy water as well. The optimized water temperature for channel catfish to grow is the warm water at the temperature of about 85° F (29.4° C) (Wellborn, 1998).

Sex determination

There are various mechanisms for sex determination in the natural world. The sex determination mechanism can be categorized into either gonochorism or hermaphroditism. Hermaphroditism is rare, while, gonochorism is more common. In gonochoristic species, different sexes could cause individual phenotypic differences. The differences in sexual characteristics can be very small (only different in the gonads) or quite significant (in the internal and external phenotypes, behavior, and metabolism). Primary sex determining types include genetic sex determination (GSD) and environmental sex determination (ESD). The most extensive sex-determining mechanism is GSD. For most mammals, the sex is determined by genetics, such as human. However, for some animals, the question of whether the progeny will be a male or a female has nothing to do with the genes at all. In this case, external factors such as temperature, pH, behavior, hormones, physiological can influence the fate of an organism and the process of gonad differentiation (eg. alligators and most turtles) (Baroiller et al., 2009). Chromosomal sex determination (CSD) is the main mode of GSD in most mammalian species. In this case, sex is determined by a primary switch located on one or both members of a well-differentiated sex chromosomal pair. The other type of GSD is polygenic sex determination (PGSD), in which the genes with strong influence on sex determination and/or gonad differentiation are distributed throughout the genome and the combination of their alleles

determines the sex of an individual (Liew et al., 2012; Nakamura et al., 1998). The mechanism is more complex in some fish species that have a sex change. The question of whether the offspring will be a male or a female isn't settled until even later in life. In many polygynous species, if the dominant female on the group dies, the largest most dominant male will then quickly become female and take her place. All the other males will move up one rank in the hierarchy.

Early in the 1900s, sex chromosomes were discovered, which was considered to be the first major breakthrough in understanding sex determination. After analyses of various species over the years, scientists discovered that chromosomal differences are primarily responsible for sex determination in most animals. Most insects, like the majority of animals, have dimorphic sex chromosomes that can be distinguished cytologically. However, sex determination mechanisms in insects are considerably diverse (Saccone et al., 2002). The sex determination mechanism in the fruit fly *Drosophila melanogaster* is relatively rare. It was reported that the ratio (X:A) of the number of X chromosomes to the number of sets of autosomes determined its sex (Cline and Meyer, 1996). In eutherian mammals, *Sry* gene on the Y chromosome is responsible for sex determination (Koopman et al., 1991; Sinclair et al., 1990). The embryo can't develop into a male without the presence of *Sry*, the master regulation for sex determination. Working with birds and chickens, sex is determined by chromosomes known as the Z and W, and females are the heterogametic sex. In chickens, a strong candidate avian sex-determinant DMRT1 is required for testis determination (Smith et al., 2009). While in *Xenopus laevis* (ZZ/ZW-type sex determination), sex determination is triggered by the ovary-determining gene, DM-W (Yoshimoto and Ito, 2011; Yoshimoto et al., 2008).

Studies on the sex determination in fish

In fish, sex determination system exhibits high diversity. Generally, the sex determination mechanisms include genetic and environmental regulation. A great variety of sex determination mechanisms exists in fish among individuals within a population, contrary to the apparent conservation of the gene differentiation network in vertebrates (Graves and Peichel, 2010). In genetic systems, certain components, or combinations of components may become dominant in influencing the direction of sex determination, while environmental factors have little influence. Some gene(s) may control the fish sex and direct ovarian and testicular development. The strength of the genetic factors the fish receives from its parents would determine the sex of a particular individual (Devlin and Nagahama, 2002). In environmental regulation, the environment the fish live in affects the embryo development crucially. In most cases, this influence is associated with the change of temperature (TSD). For example, normally, the sex of channel catfish is determined genetically by an XY system, but high temperature extremes applied during the critical period for sex determination result in female-skewed sex ratios, which indicates influence by environmental factors as well (Patino et al., 1996). This observation suggests that the sex of channel catfish is a combination of both GSD and ESD because both genetic and temperature-dependent mechanisms functions in determine the sex. Similar phenomenon has been seen more widespread in fishes than previously believed. The reason is that temperature fluctuations can dramatically influence the structure and function of proteins and other macromolecules, which encountered the possibility of the alteration of the gonad

development. In some cases, utilization of these influences could be a strategy to improve reproductive success, whereas in others, the effects on sex determination may not occur naturally, and may arise from disruptions of normal sex-determination processes under extreme environmental conditions (Devlin and Nagahama, 2002; Nakamura et al., 1998).

The sex chromosome types in fish sex determination are varied. The male heterogamety (XY) is the most common sex determination mechanism as reported in fish. For example, medaka (*Oryzias latipes*) (Matsuda et al., 2002), channel catfish (*Ictalurus punctatus*) (Patino et al., 1996), and rainbow trout (*Oncorhynchus mykiss*) have a XY system (Yano et al., 2012). However, many species have female heterogamety (ZW). In this case, ZZ individuals are male and ZW are female. Fish such as the turbot (*Scophthalmus maximus*) (Martinez et al., 2009), half-smooth tongue sole (*Cynoglossus semilaevis*) (Shao et al., 2010) have a ZW sex-determination system.

Extensive research efforts have been made to study GSD in fish species, especially with the rapid development of modern molecular genetics and genomics techniques. Up to date, five master sex-determining genes among fish species have been identified. Together with the sex-determining genes identified in mammals, chickens and *Xenopus laevis*, here, we provide a brief overview of the research on sex-determining genes that can be found (Table 1).

Table 1 Previously reported candidate “master” sex-determining genes in vertebrates

Species	Sex-determining genes	References
Human	Sry	Sinclair et al., 1990
Mice	Sry	Koopman et al., 1991
Chicken	DMRT1	Smith et al., 2009
African clawed frog	DM-W	Yoshimoto et al., 2008
Medaka, <i>Oryzias latipes</i>	DMY/dmrt1bY	Matsuda et al., 2002
Fugu	Amhr2	Kamiya et al., 2012
Patagonian pejerrey	Amhy	Hattori et al., 2012
Medaka, <i>Oryzias luzonensis</i>	Gsdf ^Y	Myosho et al., 2012
Rainbow trout	sdY	Yano et al., 2012

Besides, progress has been made in identifying several candidate sex-determining genes and loci in other fish species, such as zebrafish, tilapia, stickleback, half-smooth tongue sole, and guppy (Bradley et al., 2011; Kikuchi et al., 2007; Peichel et al., 2004; Shao et al., 2010; Shirak et al., 2006; Tripathi et al., 2009). Below, we focus on the events related to the studies on fish sex determination.

Medaka (*Oryzias latipes*)

Freshwater fish species medaka possesses a male heterogametic XY system. The first sex-determining master gene in non-mammalian vertebrate was identified in medaka (*O. latipes*) which lives in inland waters of East Asia (Matsuda et al., 2002). The master sex-determining gene Dmy/dmrt1bY (DM domain gene on the Y chromosome/doublesex and mab-3 related transcription factor 1b on the Y chromosome) is a duplicated copy of the autosomal Dmrt1 gene.

The expression of Dmy is the necessity and sufficiency for the testis development in medaka as well.

In the study of identification of the sex-determining gene in medaka, recombinant breakpoint analysis was firstly employed to restrict the sex-determining region to a 530-kilobase (kb) stretch of the Y chromosome. Deletion analysis of the Y chromosome of a congenic XY female further shortened the region to 250 kb. Shotgun sequencing of this region predicted 27 genes. Three of these genes were expressed during sexual differentiation. A single candidate gene was identified at the top of the sex determination cascade, and it's the only Y specific DM-related PG17 by positional cloning in the sex-determining region of the Y chromosome, named DMY. This gene was originated by duplication of a well-conserved downstream gene *dmrt1* on the Y chromosome (Nanda et al., 2002). In addition, functional experiments including two naturally occurring mutations were performed to verify DMY's critical role in male development. The first heritable mutant with a single insertion in exon 3 and the subsequent truncation of DMY resulted in all XY female offspring. Similarly, the second XY mutant female showed reduced DMY expression with a high proportion of XY female offspring. During normal development, DMY is expressed only in somatic cells in the testis. These findings strongly suggest that the sex-specific DMY is required for testicular development and is a prime candidate for the medaka sex-determining gene. Continuous study in 2007 showed that a 117-kb genomic DNA fragment that carries DMY is able to induce testis differentiation and subsequent male development in XX (genetically female) medaka (Matsuda et al., 2007). In addition, overexpression of DMY cDNA under the control of the CMV promoter also caused XX sex

reversal. These results demonstrate that DMY is sufficient for male development in medaka and suggest that the functional difference between the X and Y chromosomes in medaka is a single gene. In conclusion, DMY is the sex-determining gene in medaka. The discovery of Dmy was a framework for understanding genetic sex determination in the following fish species. However, due to the extensive diversity of this sex-determining gene, DMY doesn't show its sex-determining feature even in the closely related species (Kondo et al., 2006; Kondo et al., 2003).

Medaka (*Oryzias. luzonensis*)

Since DMY is not the master sex-determining gene in all medaka fish, additional work needs to be done to identify SD gene in other medaka fish. Recently, the discovery of a novel SD gene Gsdf^Y (gonadal soma derived growth factor) has been reported by using a genetic mapping approach. It was considered to be a replacement for Dmy because Dmy was lost during a transposition event (Myosho et al., 2012).

In attempt to identify the SD gene, linkage analysis was performed and two male recombinants for the SD region of *O. luzonensis* were obtained. A BAC library of one XX fish and a fosmid library of one super male YY fish were constructed respectively and physical maps of the SD region of the X and Y chromosomes were made. SD region for both the X and Y chromosomes exhibits high sequence identity with no large deletions or insertions, and covers 9 predicted genes identified by Genscan program. RT-PCR was performed to examine whether the predicted genes are expressed during sexual differentiation. The results show that only one gene, PG5 expressed higher in XY embryos than in XX embryos. Chromosome walking was carried

out to obtain the full-length mRNA sequence of PG5. Phylogenetic analysis found it shown in the same clade as Gsdf. Comparing Gsdf on the X chromosome (Gsdf^X) with that on the Y chromosome (Gsdf^Y), 12 base substitutions were found in the full-length mRNA, including two synonymous substitutions in the ORF; however, the amino acid sequences are the same. Real-time PCR was employed to examine the expression profiles of Gsdf^X and Gsdf^Y. Expression of Gsdf was higher in XY embryos than in XX embryos from 2 days before hatching (dbh) to 0 day after hatching. Overexpression of Gsdf^Y induced XX individuals into fertile XX male. Luciferase assays demonstrated that the upstream sequence of Gsdf^Y contributes to the male-specific high expression. Gsdf is downstream of Dmy in the sex-determining cascade of *O. latipes*, suggesting that emergence of the Dmy-independent Gsdf allele led to the appearance of this novel sex-determining gene in *O. luzonensis*.

Patagonian pejerrey (*Odontesthes hatcheri*)

Patagonian pejerrey, the freshwater native fish of Argentina, belongs to a group of Atherinopsid fish (silversides) widely spread in inland and marine waters of the southern South America. With a tasty flesh of excellent quality, Patagonian pejerrey would probably become a candidate for aquaculture (Hualde et al., 2011).

In Patagonian pejerrey, Amhy gene, which is a male-specific, functional duplicated copy of the anti-Mullerian hormone gene (amh) was suggested to be involved in the sex determination in this species (Hattori et al., 2012; Kikuchi and Hamaguchi, 2013). Amh is a member of the TGF- β superfamily and a well-characterized hormone in mammals and plays an important role in

the development and maintenance of reproductive organs (Drummond, 2005; Fan et al., 2012; Mishina et al., 1996). By using molecular cloning, sequencing, genome walking, and fluorescence *in situ* hybridization (FISH), the male-specific gene *amhy* was amplified and sequenced. Compared with its homolog *amha*, a 557-bp *amhy*-specific insertion in the third intron of the gene was identified within un-transcribed intragenic regions as the main structural difference. Expression analysis was performed and during embryonic and early larval development of both sexes and *amhy* was detected to be expressed as the first signs at the critical testicular differentiation stages in XY males. The targeted inactivation of *amhy* in males using morpholino-mediated knockdown induces ovarian differentiation in XY fish, with a sex reversal rate of 22%. Taken together, *amhy* is probably a sex-determining hormonal gene in Patagonian pejerrey.

Rainbow trout (*O. mykiss*)

Rainbow trout is regarded as an important model of early vertebrate sex chromosome evolution. It possesses a simple XY monofactorial system with a strict GSD type. Recently, Yano et al (2012) proposed the identification of a novel vertebrate master SD gene *sdY* (Yano et al., 2012). This gene, which has a similarity with an immune-related gene interferon regulatory factor 9, was found to be specifically expressed in the differentiating testis with homology to a rainbow trout Y chromosome sequence by next-generation sequencing (NGS). The expression of *sdY* gene was detected during early testicular development using *in situ* hybridization. The gene expression level of *sdY* reached the peak at approximately 40 to 60 days post fertilization (dpf)

only in male differentiating gonads, and this high expression was sustained until 90 dpf by using quantitative PCR (qPCR). In order to locate sdY to the rainbow trout genetic map, markers derived from the coding portion of sdY were employed. The results showed that sdY located a few kilobases away from the polymorphic Y-linked OmyY1 marker in rainbow trout and was strictly colocalized with the SEX locus on rainbow trout linkage group 01 (Brunelli et al., 2010; Brunelli et al., 2008). To prove that the function of sdY is consistent with a role as a master sex-determining gene, whether the overexpression of sdY induced testicular differentiation in rainbow trout genetic females (XX) was investigated. In addition, the targeted inactivation of sdY was produced in genetic males (XY) using zinc-finger nucleases (ZFN). Taken together, the conclusion was reached that sdY is a novel vertebrate master sex-determining gene.

Fugu

Takifugu rubripes is one of the most economically important food fish in East Asia. The whole genome of fugu has been sequenced and completed, but the sex determination mechanism has not been figured out until the recent discovery of a SNP (C/G) in the anti-Mullerian hormone receptor type II (Amhr2) gene (Kamiya et al., 2012). This single locus on chromosome 19 is the only polymorphism associated with phenotypic sex (Kai et al., 2005; Kikuchi et al., 2007). A combination of the two alleles of Amhr2 was predicted to be responsible for the sex determination in fugu. Males are heterozygous (His/Asp384) in the kinase domain, while females are homozygous (His/His384). Examination of this locus in the fugu genome database suggested that the candidate region includes at least four scaffolds containing ~300 potential protein-coding

genes. Marker density across the whole genome was increased to obtain a contiguous physical map of the sex chromosome. Family-based genetic mapping and association mapping were constructed to precisely pinpoint the SD gene utilizing ancestral recombination in a wild population of fugu. SNPs and other polymorphisms were screened by sequencing the entire SD region. The SNP locus was identified with 100% accuracy of all heterozygous males and all homozygous females tested at this position.

More interestingly, the sex-determining SNP in fugu locates in a region with no signs of recombination suppression. This is quite different from the sex-determining genes identified so far such as therian Sry (Graves, 2006) and medaka Dmy (Kondo et al., 2006). Such a phenomenon would lead us to think of the possibility for a more common expression of the undifferentiated X-Y chromosomes in vertebrates (Kamiya et al., 2012).

Zebrafish

Zebrafish has become an increasingly important vertebrate model organism. However, there is a lack of study on the mode of sex determination in zebrafish. Various conclusions have been generated and the debate over the putative sex chromosome is on (Anderson et al., 2012; Bradley et al., 2011; Liew et al., 2012). Zebrafish sex determination system was proposed to be a polygenic sex determination system with influences from the environment (Liew et al., 2012). Classical breeding experiments were performed together with large-scale genomic analyses. No sign of a chromosomal sex determination system was detected in zebrafish genetically. In the latest publication of the zebrafish reference genome sequence, an enrichment of zebrafish-

specific genes on chromosome 4 and chromosomal regions that influence sex determination was identified, which is definitely a strong support to propose that this chromosome may be, might have been, or may be becoming a sex chromosome in this particular population (Howe et al., 2013). In addition, other separate genomic regions have been identified as well. A notable difference reported a genome-wide linkage study of sex determination in zebrafish using a novel SNP genetic map (Bradley et al., 2011). The linkage results demonstrate that sex determination is a complex trait in zebrafish, not employing sex chromosomes. Two loci on zebrafish chromosome 5 and 16 were described to contribute to sex determination significantly, rather than as an XY or ZW genetic system. Each of these loci contains a prominent candidate gene with a conserved role in sex determination. However, additional important components of sex determination including interacting environmental cues and complex genetic factors remain to be elucidated in zebrafish. Another event during the process in understanding the genetic basis of zebrafish sex determination is the report that chr-4 has functional and structural properties expected of a sex chromosome (Anderson et al., 2012). Again, zebrafish was confirmed to possess a polygenic sex determining system with an important determinant occupying a short 2 Mb region at the tip of chr-4q, the only arm in the zebrafish genome that shares many features of sex chromosomes.

Stickleback

Stickleback fish are an excellent system to analyze the genetic and molecular mechanisms that underlie the evolution of sex determination and sex chromosomes. Genome-

wide linkage mapping studies were carried out on three-spine stickleback, and identified the location of the SD region in the distal section of LG19 (Peichel et al., 2004). A XY sex-determining system was suggested in three-spine stickleback even though the sex chromosomes are invisible cytogenetically. Additionally, suppression of recombination was observed in males in this SD region. There is an accumulation of repeat sequences and duplications detected in the 250 kb region through BAC clone analysis. A significant difference has been observed in the SD region or system in different stickleback species. Molecular marker analysis revealed that the nine-spine stickleback (*Pungitius pungitius*) exhibits similar XY mechanism associated with LG12. As in *Gasterosteus weathlandi*, a combination two SD chromosomes LG12 and LG19 was presented (Ross et al., 2009). Four-spine stickleback (*Apeltes quadracus*) and brook stickleback (*Culaea inconstans*) showed ZW system. ZW pair in these species is not homologous to other stickleback species by hybridization with LG12 and LG 19 probes (Shapiro et al., 2009).

Tilapia

Over 70 species of Cichlids are included in the wide variety of tilapias. Previous study identified that QTL on both LG1 and LG3 and other secondary components were associated with sex (Shirak et al., 2006). Properties of these two chromosomes confirmed their association with sex as well because of the existence of a large amount of transposon elements and an apparent decrease of recombination. SD system differs in different tilapias according to Cnaani et al. (2008) (Cnaani et al., 2008). *O. niloticus* and *Tilapia zillii* are tilapias with a XY system and SD locus locates in LG1; however, the fixation of SD region in *O. karongae* and *T. mariae* suggests

a ZW system with the SD locua in LG3. Even though GSD has been reported to play a crucial role in determine the sex of tilapias, TSD also was described to impose significant influence on sex determination in tilapias (Baroiller et al., 2009). Hence, it will be rewarding to understand the complicated signatures in so many species in the diverse group of tilapias.

Catfish

Channel catfish normally possess a relatively simple genetic sex determination system (XX/XY male heterogametic system) (Davis et al., 1990; Tiersch et al., 1990; Tiersch et al., 1992), but high temperature extremes applied during the critical period for the sex differentiation could cause female-skewed sex ratios. This suggested that environmental factors, such as temperature, can influence the sex determination of catfish as well (Patino et al., 1996).

Previous molecular genetics analysis was conducted with genes such as *Sry* (mammalian sex-determining gene) and its closely linked gene *Zfy* in channel catfish. The results showed that even though these genes existed in the catfish genome, they were not seen to express in a sex-specific pattern (Tiersch et al., 1992). An isozyme locus GPI-B that was very close to the centromere was found on the sex chromosomes in channel catfish. This locus was located approximately 16 map units away from the sex-determination locus and on the other side of the centromere (Liu et al. 1996). Seven microsatellite loci closely linked with the sex were identified when constructing the microsatellite-based linkage map (Waldbieser et al., 2001). However, these loci were restricted to a specific catfish family and can't be widely applied to other families or strains. Recently, Ninwichian et al. reported the identification of one Y-linked specific

microsatellite marker, which displayed 100% sex-typing accuracy from four different strains of channel catfish. This method has already been put into use as a rapid and efficient way to identify the sex of channel catfish in the lab. PCR amplifications of the sex-linked marker locus produced both a male-specific (Y-linked) product and an autosomal product seen in both males and females, which may lead us to consider whether or not there is a Y-specific region (Ninwichian et al., 2012b). Since the sex-related marker was identified, the chromosome that contains the sex-linked marker was designated as the sex chromosome. Then how far the marker is away from the sex-determining gene becomes a mystery.

Next-generation sequencing

The recent advent of next-generation sequencing technologies has been applied to analyse large quantities of data on genomes and transcriptomes. Huge progress has been made towards human and model organisms under various treatments using next-generation sequencing, and it is now more widely used in non-model organisms as well.

Among these powerful and rapidly evolving technologies, three main platforms for massively parallel DNA sequencing read production are in reasonably widespread use at present: the FLX pyrosequencing system from 454 Life Sciences (a Roche company), the Illumina Genome Analyzer (developed initially by Solexa), and the ABI SOLiD system (now Life Technologies). The throughput varies among the three platforms, from hundreds of thousands of reads for the 454 FLX system to hundreds of millions of reads for the Illumina Genome Analyser and ABI SOLiD systems (Mardis, 2008; Marguerat and Bahler, 2010; Shendure and Ji, 2008).

Roche/454 FLX Pyrosequencing

Roche/454 pyrosequencing sequencing was the first next-generation sequencing platform to achieve commercial introduction (Mardis, 2008; Margulies et al., 2005). The principle of this approach is briefly introduced as following (Rothberg and Leamon, 2008). The first step is the library construction, in which each bead carrying oligonucleotides complementary to the 454-specific adapter sequences is associated with a single library fragment. Each of these fragment: bead complexes is isolated into individual oil: water micelles that also contain PCR reactants. Then thermal cycling (emulsion PCR) of the micelles produces around one million copies of each DNA fragment on the surface of each bead. After amplification, the emulsion shell is broken and the clonally amplified single molecule beads are ready for loading onto the fibre-optic PicoTiter Device for sequencing. The PicoTiter Plate is loaded with one fragment carrying bead per well and smaller beads with the enzymes that catalyze the downstream pyrosequencing reaction steps. The CCD camera that records the light emitted at each bead captures each single signal during a number of cycles the four bases (ATGC) are sequentially washed over the PicoTiter Plate.

Illumina Sequencing

Developed from researchers from Sanger Institute, Illumina's sequencing by synthesis (SBS) technology is the most successful and widely-adopted next-generation sequencing platform worldwide (http://www.illumina.com/technology/sequencing_technology.ilmn).

Currently, the main systems in Illumina are HiSeq and MiSeq. Mostly important technique is the application of TruSeq technology to use labeled nucleotides with reversible terminators to sequence a single base at a time. This enables the massively parallel sequencing to be incorporated into growing DNA strands. The nucleic acid samples are sheared into a random library of 100-300 base-pair long fragments. Ends of the obtained DNA-fragments are repaired and a poly-A-overhang is added at the 3'-end of each strand. Adaptors necessary for amplification and sequencing are ligated to both ends of the DNA-fragments. These fragments are then size selected and purified. Bridge amplification is performed during cluster generation. Simultaneously, the clusters are being sequenced base-by-base. The fluorescence signal after each incorporation step is captured by a built-in camera, producing images of the flow cell.

ABI SOLiD Sequencing

The SOLiD platform uses a core technique of sequencing by ligation which is different from the other next-generation platforms, and uses an emulsion PCR approach with small magnetic beads to amplify the fragments for sequencing (Mardis, 2008). The steps of shearing of DNA and adaptor ligation are similar like those with 454 and Illumina platform. Each ligation step is followed by fluorescence detection. Clonal clusters are also needed in microreactors containing template, PCR reaction components, beads, and primers. SOLiD generates DNA by measuring the serial ligation of a fluorescently labeled oligonucleotide. After each ligation, the fluorescence signal is measured and then cleaved before another round of ligation takes place. SOLiD applied the two-base encoding for base-calling, enables this platform is more widely used

in genome re-sequencing and SNP discovery due to its highest accuracy. Briefly, the amplification products are transferred onto a glass surface where sequencing occurs by sequential rounds of hybridization and ligation with 16 dinucleotide combinations labeled by four different fluorescent dyes (each dye used to label four dinucleotides). Each position can be read twice and the color is determined by two successive ligation reactions (Morozova and Marra, 2008).

In addition to the above three major sequencing platforms, the semiconductor sequencing technology Ion Torrent has pioneered an entirely new approach to DNA sequencing. Ion Torrent employs “sequencing by synthesis” method based on the detection of hydrogen ions that are released during the polymerization of DNA. Unlike previous next-generation platforms, Ion Torrent doesn’t require fluorescence or camera scanning, which significantly reduces the cost and increases the speed (Liu et al., 2012a). Besides, MiSeq by Illumina is gaining its popularity due to its significant improvement in read length and assembly quality.

The third generation sequencing technology, such as Pacific Bioscience, Nanopore sequencing, and Helicos, has great potential to be widely applied in the sequencing field with its unique characteristics. The most attractive feature imported by Pacific Bioscience is Single-molecule real-time (SMRT). The whole sequencing process doesn’t require PCR amplification of the nucleic acid fragments, which is definitely time-saving. Moreover, DNA polymerase is observed working in real-time, which suggests that the signal is monitored during the enzymatic reaction of adding nucleotide in the complementary strand, no matter whether it is fluorescent (PacBio) or electric current (Nanopore) (Liu et al., 2012a). Taken together, the advent of so

many novel sequencing platforms did revolutionize the sequencing landscape and become a dominant technique driven by the rapidly dropping cost. Here, we list the comparison of the sequencing statistics in various mainstream platforms in Table 2. Decisions on which sequencing platform to be adopted should be made based on the experimental goal. Since the purpose of assembly (initial step of a genome or a transcriptome analysis) is to reconstruct sequence up to chromosome length, factors such as read length and the amount of the reads are essential because they are closely correlated with the reads overlap (Bao et al., 2011; Miller et al., 2010).

Table 2 Comparison of several important aspects of the sequencing statistics in various mainstream platforms. Run time, throughput and cost may vary significantly depending on sequencing instruments, experimental design, core facility discounts, etc.

Platform	Roche 454	Illumina	ABI SOLiD	Pac Bio	Ion Torrent
Principle	Sequencing by Synthesis	Sequencing by Synthesis	Sequencing by Ligation	Single molecule sequencing	Semiconductor Sequencing
Library construction	emPCR/ Clonal cluster	Bridge Amplification/ Clonal Cluster	emPCR/ Clonal cluster	No amplification	No amplification
Sequencing chemistry	Pyrosequencing	Reversible Terminator	Ligase activity	Real-time	Chemical to digital
Read length (bp)	~400	35-150	50-75	~3000	100-200
Throughput/ run	1 million	Up to 2 billion	Up to 2 billion	20 billion	60-80 million
Pros	Longer reads; Faster run time	Most widely used	Highest accuracy	Very fast run time; longest read length	User-friendly; fast run time
Cons	Higher cost	Short read length	Long run time	High error rate	Lower throughput

Assembly strategies and tools

Sequencing assembly refers to aligning and merging short reads into contigs in order to reconstruct the original sequence. It is always challenging due to the existence of the repetitive elements and the requirement of intensive computation resources (Ji et al., 2011). General idea for assembly is to construct an overlap graph after comparing the reads to each other using the assemblers in the first overlap discovery stage. Then analysis of the overlap graph is needed to traverse the graph into appropriate paths in the layout stage. Finally, consensus sequences (contigs) would be generated through multiple sequence alignment (Bao et al., 2011).

Major approaches for assembly encompass comparative assembly and *de novo* assembly. Comparative assembly is relatively easy to achieve in a way. Taking the advantage of the reference genome, the DNA fragments can be mapped to the reference and the mapping results would be beneficial to the inference of the genome (Pop, 2009; Pop et al., 2004). For the genome of an organism has not been sequenced and assembled, *de novo* assembly from next generation sequencing data should be applied to bring the short sequence reads from random positions along a target molecule. To accomplish the *de novo* assembly, two main strategies are put into use. One algorithm is the overlap-layout-consensus (OLC) approach and the other is designed based on a *de Bruijn* graph (DBG) (Cahais et al., 2012; Robertson et al., 2010). DBG-based assemblers stand out and being more widely adopted between the two because it was reported that DBG-based assemblers usually provide more accurate assembly results (Paszkiwicz and Studholme, 2010).

Various assemblers were designed based on the *de Bruijn* graph approach, such as Velvet (Zerbino and Birney, 2008; Zerbino et al., 2009), ABySS (Simpson et al., 2009), AllPaths (Butler et al., 2008; MacCallum et al., 2009), SOAPdenovo (Li et al., 2010), etc. Velvet, a full implementation of DBG assembly, employs the graph simplification and compresses the graph without loss of information. It has an error-avoidance read filter and applies a series of heuristics that reduce graph complexity. Velvet targets *de novo* assembly from short reads with paired ends from the Illumina platform. However, it might not be suitable for assembling very large genomes because of the memory requirements. ABySS is a scalable assembly software for Illumina short reads and paired end reads. It could distribute the K-mer graph, and the graph computations, across a compute grid whose combined memory is quite large. Even though ABySS does not build scaffolds, it is very competitive to generate contigs with perfect sequence overlap. With the ability to be applied for large genome assembly, AllPaths exploits paired-end short reads from the Illumina platform and simplifies its K-mer graph initially based on reads and overlaps. The SOAPdenovo program could deal with large genome sequences with high quality. Using the DBG method, SOAP builds contigs from all the sequencing reads. Then making use of the paired reads, SOAP combines the contig consensus sequences to build scaffolds (Miller et al., 2010). Overall, each assembly software has its unique characteristic. When we choose which one to use for our own data, we need to put several factors into consideration and make comparisons among the results produced. The contiguity and the accuracy of contigs or scaffolds are crucial indexes for evaluating the assembly quality. The contiguity mentions lengths of contigs or scaffolds, such

as the total length, the average length and the longest length, etc. The accuracy indicates the error rates for the assembly (Ji et al., 2011).

CHAPTER II. OVERVIEW AND RESEARCH OBJECTIVES

Catfish, one of the lower teleosts, is the dominant aquaculture species in the United States taking up more than 60% of U.S. aquaculture production (Hanson and Sites, 2013). Channel catfish, as a good research model, are native to the Mississippi River drainage and their range has expanded to most regions of North America. Male channel catfish usually grow faster than females in the wild (Kelly 2004). In addition, sex is considered to be one of the most fundamental features in the life. Due to economic aspects as well, it's worthwhile to study and illustrate the mechanism of catfish sex determination.

Sex determination mechanism exhibits remarkable plasticity and diversity in teleost. Channel catfish normally possess a relatively simple genetic sex determination system (XX/XY male heterogametic system) (Davis et al., 1990; Tiersch et al., 1990; Tiersch et al., 1992), but high temperature extremes applied during the critical period for the sex differentiation could cause female-skewed sex ratios (Patino et al., 1996). Although a few studies on the sex differentiation and sex-linked markers were well documented for channel catfish (Ninwichian et al., 2012b; Waldbieser et al., 2001), little is known for its sex-determining genes.

Learning from the success in identification of the sex-determining gene in fish species, my research, as presented here, encompasses the two major approaches (using XY system as an example): 1) Identification of Y-specific sequences; 2) Identification of male-specific transcripts. In both cases, the sex-determining gene must be validated by functional analysis of the Y-specific sequences and male specific transcripts for their necessity and sufficiency for sex determination. Unlike the obviously morphological distinct sex chromosomes (X and Y) in mammals, X chromosome and Y chromosome in channel catfish are quite similar (Tiersch et al.,

1990). However, there may be the difference in the DNA content (DNA sequences) from the X and Y chromosome in channel catfish. The issue is how we can identify this region.

Analysis of genes expressed in male-specific organs has been utilized as one approach to identify male-specific transcripts. Sex-biased genes expressed predominantly or exclusively in one sex were thought to drive the phenotypic differences between males and females (Assis et al., 2012; Ellegren and Parsch, 2007). We conducted RNA-Seq of the testis tissue with a goal of profiling the genes expressed in the testes, and of exploring the possibility to identify male-biased transcripts, which might be involved in gonadogenesis, spermatogenesis, testicular determination, gametogenesis, gonad differentiation, and possibly sex determination. On the other hand, whole genome sequences of a regular male channel catfish (XY), whole genome sequences of a super male channel catfish (YY), and the genome sequences of all the BAC clones on the Y chromosome (LG4) were generated using next-generation Illumina HiSeq. We assembled and scaffolded the X and Y chromosome by means of catfish physical map (Quiniou et al., 2007; Xu et al., 2007), linkage map (Kucuktas et al., 2009; Ninwichian et al., 2012a; Waldbieser et al., 2001), and the gynogen (XX female) genome resources, followed by comparing these two. Our goal was to answer if there is a Y-specific sequence in channel catfish or not. In addition, using both linkage and physical mapping, a 203 kbp putative sex-linked region encompassing ten adjacent markers (one of them is the 100% accuracy sex-linked microsatellite marker) was identified, consisting of 16 scaffolds and harboring 10 annotated genes. Genes in this region are considered to be most likely correlated with catfish sex determination and/or sex differentiation at early embryonic development.

*** CHAPTER III. MALE-BIASED GENES IN CATFISH AS REVEALED BY
RNA-SEQ ANALYSIS OF THE TESTIS TRANSCRIPTOME**

* Published in PLOS ONE.

Abstract

Background: Catfish has a male-heterogametic (XY) sex determination system, but genes involved in gonadogenesis, spermatogenesis, testicular determination, and sex determination are poorly understood. As a first step of understanding the transcriptome of the testis, here, we conducted RNA-Seq analysis using high throughput Illumina sequencing.

Methodology/Principal Findings: A total of 269.6 million high quality reads were assembled into 193,462 contigs with a N50 length of 806 bp. Of these contigs, 67,923 contigs had hits to a set of 25,307 unigenes, including 167 unique genes that had not been previously identified in catfish. A meta-analysis of expressed genes in the testis and in the gynogen (double haploid female) allowed the identification of 5,450 genes that are preferentially expressed in the testis, providing a pool of putative male-biased genes. Gene ontology and annotation analysis suggested that many of these male-biased genes were involved in gonadogenesis, spermatogenesis, testicular determination, gametogenesis, gonad differentiation, and possibly sex determination.

Conclusion/Significance: We provide the first transcriptome-level analysis of the catfish testis. Our analysis would lay the basis for sequential follow-up studies of genes involved in sex determination and differentiation in catfish.

Keywords: catfish, fish, sex determination, sex-biased expression, RNA-Seq, transcriptome

Introduction

Sex is one of the most fundamental features of life. In teleost, sex determination mechanism exhibits extraordinary plasticity and diversity during evolution. Fish represent over 50% of all vertebrates with over 32,400 species. The vast majority of fish species are gonochoristic although some fish are hermaphroditic. The latter is unique in fish among vertebrates (de Mitcheson and Liu, 2008; Kobayashi et al., 2013). With gonochoristic fishes, the sex could be either genetic sex determination (GSD) or environmental sex determination (ESD), and sometimes a combination of both. In most cases among fishes, sex is determined by GSD, but environmental factors, especially temperature, can influence sex determination (Devlin and Nagahama, 2002).

The most common sex determination system in fish is the XY system, i.e., male heterogamety. However, in many species, the sex determination system is the ZW system, i.e., female heterogamety. For instance, medaka (*Oryzias latipes*) (Matsuda et al., 2002) and rainbow trout (*Oncorhynchus mykiss*) (Yano et al., 2012) have XY sex determination system, while turbot (*Scophthalmus maximus*) (Martinez et al., 2009) and half-smooth tongue sole (*Cynoglossus semilaevis*) (Shao et al., 2010) have ZW sex determination system. Even in closely related species, the sex determination system may vary. For example, in tilapia, it is believed that both XY and ZW systems are functional depending on the species (Devlin and Nagahama, 2002). Due to the unrivaled diversity, the study of sex determination in fish becomes a very intriguing research area. Such studies have demonstrated unprecedented diversity. However, to date, sex

determination mechanisms have been elucidated only in five fish species including medaka (*O. latipes*) (Matsuda et al., 2002), rainbow trout (Yano et al., 2012), fugu (Kamiya et al., 2012), Patagonian pejerrey (*Odontesthes hatcheri*) (Hattori et al., 2012) and another species of medaka (*O. luzonensis*) (Myosho et al., 2012). In medaka (*O. latipes*), DMY gene was found to be the sex determination gene, which is a duplicated gene of Dmrt1 on the Y-chromosome. In Fugu, there was no Y-specific gene that controls sex. Instead, a single nucleotide polymorphism within the anti-Mullerian hormone receptor type II (*Amhr2*) gene was found to control the sex (Kamiya et al., 2012). In rainbow trout, SDY gene, a truncated immune related gene interferon regulatory factor 9 like gene residing on the Y-chromosome, was found to be the sex control gene. With Patagonian pejerrey, *Amhy* gene, which is a male-specific, functional duplicated copy of the anti-Mullerian hormone gene (*amh*) was suggested as a strong candidate of the master sex-determining gene in this species (Hattori et al., 2012). In addition to these where the sex determination gene has been identified, active research is underway with a number of aquaculture fish species because of potential applications of sex determination mechanisms for aquaculture production (Bradley et al., 2011; Kikuchi et al., 2007; Peichel et al., 2004; Shao et al., 2010; Shirak et al., 2006; Tripathi et al., 2009).

Channel catfish (*Ictalurus punctatus*) is the most important aquaculture species in the United States, and also an excellent model organism for teleost genetics and genomics studies (Liu, 2003). Sex in catfish is mainly determined by the genetic sex determination (GSD) system, with the co-existence and interaction with temperature-dependent sex determination (TSD) (Patino et al., 1996). Genetically, catfish possess XY male heterogametic system. Although the

sex differentiation was well documented for channel catfish (Green and Kelly, 2009; Patino et al., 1996), little is known for its sex-determining genes.

Two major approaches have been used for studies of sex-determining genes in teleost fish (using XY system as an example): 1) Identification of Y-specific sequences; 2) Identification of male-specific transcripts. In both cases, the sex-determining gene must be validated by functional analysis of the Y-specific sequences and male specific transcripts for their necessity and sufficiency for sex determination. Apparently, when the Y-chromosome is highly divergent from the X chromosome, the first approach is quite advantageous. However, no significant difference in DNA content was detected from male and female cells in channel catfish, which is consistent with the reports in many other fish species in which the karyotype of X chromosome and Y chromosome is very similar (Tiersch et al., 1990). In such cases, male and female genomes carry identical or almost identical DNA. It is possible that the subtle differences of a very small portion of genes located on sex-specific chromosomes or sex-determining region are responsible for the sex determination (Ellegren and Parsch, 2007; Gallach et al., 2011; Leder et al., 2010), and discovery of the Y-specific sequences may be difficult. In the extreme situation under this case like in the case of *Fugu*, there are no Y-specific sequences or Y-specific transcripts. Instead, only a single SNP is responsible for sex determination. Therefore, the study of sex determination gene can then be extremely difficult. Apparently, analysis of male-specific transcripts could provide some advantage than analysis for Y-specific sequences in the absence of a whole genome sequence. One possibility does exist, i.e., some genes may be expressed only in the males, but not in the females. Obviously, testis is a male organ and therefore transcriptome analysis using

testis tissue could be of interest for the potential detection of male-biased transcripts, which may or may not involve sex determination gene(s).

Nevertheless, analysis of genes expressed in male-specific organs has been utilized as one of the approaches to identify male-specific transcripts. Sex-biased genes expressed predominantly or exclusively in one sex were thought to drive the phenotypic differences between males and females (Assis et al., 2012; Ellegren and Parsch, 2007). It was reported that in adult zebrafish the difference of the key transcript abundance between males and females may cause the phenotypic sexual dimorphism mediated by sex differences in gene expression, and these differences likely play a major role in phenotypic sexual dimorphism (Small et al., 2009). A large fraction of the expression divergence between sexes occurs in the gonads (Hale et al., 2011), and the sex-biased genes are expected to exhibit a higher possibility to be male-specific genes involved in sex determination (Parisi et al., 2003).

In catfish, we have previously made major efforts for transcriptome analysis. Early efforts used EST analysis (Cao et al., 2001; Ju et al., 2000; Karsi et al., 2002; Kocabas et al., 2002; Li et al., 2007; Wang et al., 2010), and such collective efforts allowed identification of around 500,000 ESTs in catfish. Recently, the application of the nextgen sequencing technologies allowed identification of over 25,000 genes in catfish through RNA-Seq (Liu et al., 2012b; Liu et al., 2011). Such efforts have led to sequencing as well as assembly of over 14,000 full length cDNAs (Chen et al., 2010; Liu et al., 2012b). However, in all such transcriptome analysis, no testis tissues were used, and therefore, genes specifically or preferentially expressed in the testis could have been missed. As a part of the catfish genome program to provide expression evidence for

gene models for genome assembly and annotation, the major objective of this study was to generate additional transcriptome information from the testis tissue. In this study, we took advantage of the high throughput next generation sequencing and conducted RNA-Seq of the testis tissue with a goal of profiling the genes expressed in the testis. Because testis is a male specific organ, our additional interest is to identify male-biased transcripts, which could contribute to our ongoing efforts to elucidate sex determination mechanisms in catfish.

Results

Sequencing and assembly of short expressed reads from catfish testis

A total of 294.6 million reads of 100 bp long were generated. Ambiguous nucleotides, low-quality sequences (quality scores < 20) and short reads (length < 30 bp) were removed, and the remaining high-quality reads (91.5 %) were carried forward for transcriptome assembly and analysis (Table 1).

Table 1. Summary of RNA-Seq of the testis transcriptome

Name	Number
Number of reads from raw data	294,618,596
Average read length (bp)	100
Number of reads after trimming	269,572,695
Percentage retained	91.5%
Average read length after trimming (bp)	93.9

ABYSS & Trans-ABYSS was used for the assembly of the RNA-Seq short reads because they provided a superior assembly when compared with Velvet and CLC Genomics Workbench (Li et al., 2012). Use of Trans-ABYSS to merge ABYSS multi-k-assembled contigs resulted in an initial approximately 6.0 million contigs with average length of 129.5 bp and N50 size of 163 bp. A total of 340,000 contigs with lengths greater than 200 bp were carried forward for additional analysis. CAP3 was employed to remove redundancy generated during multi-k assemblies. Following CAP3, 193,462 contigs with an average length 570.7 bp and N50 size of 806 bp were designated as the channel catfish testis transcriptome in the following steps of analysis (Table 2).

Table 2. Assembly statistics of the testes transcriptome using Trans-ABYSS

Assembly Metrics	Number
No. of assembled contigs from ABYSS & TransABYSS	5,977,085
No. of contigs after removing redundancy	193,462
Average length of final testes transcriptome (bp)	570.7
Final N50 (bp)	806
Total number of known unigenes	25,307
Total number of newly identified catfish transcripts	335

Gene identification and annotation of channel catfish testis transcriptome

Gene identification of the assembled contigs was performed using BLASTX searches against the reference proteins available at NCBI zebrafish RefSeq and Uniprot protein databases to annotate the channel catfish testis transcriptome. After annotation, approximately 68,000

assembled contigs had hits in non-redundant database. A total of 55,521 assembled contigs had significant ($E\text{-value} \leq 1e\text{-}10$) hits to the zebrafish RefSeq database, corresponding to 17,180 unique proteins. In searches against the Uniprot database, 47,383 catfish contigs had significant hits, accounting for a total of 21,877 unique proteins. Cumulatively, a total of 67,923 catfish contigs had at least one significant hit against the two databases, allowing 25,307 unigenes to be identified in channel catfish testis transcriptome (Table S1).

Unique gene-coding contigs from the channel catfish testis reference assembly were then used as inputs to perform gene ontology (GO) annotation by Blast2GO (Conesa et al., 2005). A total of 28 GO terms including 13 (46.4%) cellular component terms, 5 (17.9%) molecular functions terms and 10 (35.7%) biological process terms were assigned to 25,307 unique gene matches. The percentages of annotated catfish sequences assigned to GO terms are shown in Figure 1. Analysis of level 2 GO term distribution showed that cell (GO:0005623), binding (GO:0005488), and cellular process (GO:0009987) were the most common annotation terms in the three GO categories (Figure 1).

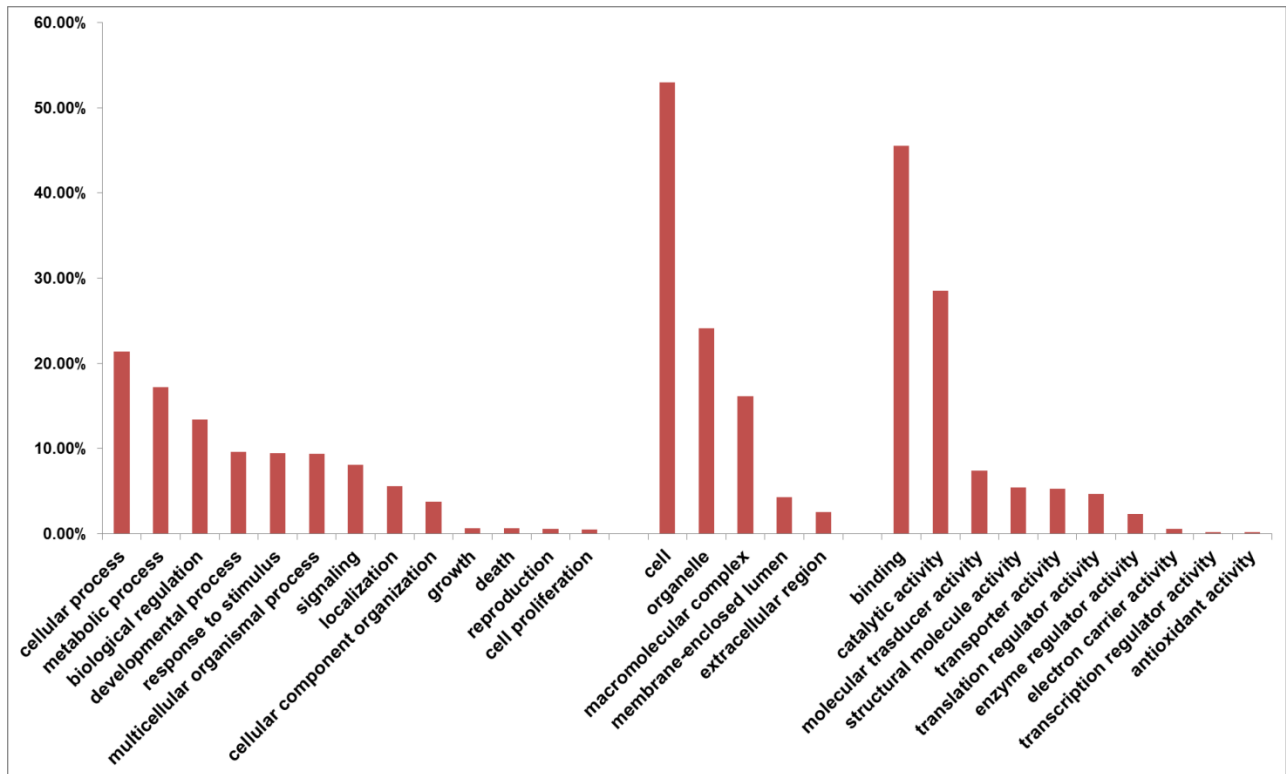


Figure 1. GO term classifications of channel catfish testes transcriptome. GO annotations were based on zebrafish RefSeq and GO-terms were processed by Blast2Go and categorized at level 2 into three major functional categories (biological process, cellular component, and molecular function).

Newly identified genes in catfish

To identify transcripts that were newly identified from the testis, all assembled contigs from testis transcriptome were used as queries to perform BLASTN searches against the existing catfish transcriptome. A total of 335 unique novel contigs from testis transcriptome had no blast hits against the existing catfish transcriptome assemblies. As more than one contigs had hits to a

single unique gene, these 335 contigs corresponded to 167 putative unique novel genes when searched against the NR database using BLASTX (Table S2), suggesting that 167 genes were newly identified from the catfish testis.

Identification and distribution of the putative male-biased genes

In order to identify male-biased genes, i.e., genes that are preferentially expressed in the male, a meta-analysis of RNA-Seq data was conducted to compare expression ratios in the testis (male) and in the gynogen (female). The gynogen was a doubled haploid fish that are homozygous without the Y-chromosome (Waldbieser et al., 2010). *In silico* mapping of RNA-Seq reads allowed 81% high quality sequencing reads of the testis RNA-Seq to be mapped onto the testis transcriptome, while 71% high quality sequencing reads from the gynogen catfish RNA-Seq (Liu et al., 2012b) were mapped onto the testis transcriptome. Using a 5-fold ratio cut-off (expressed five times more in male than in the female), a total of 5,450 genes were found to exhibit preferential expression in the testis. Of these, 637 genes were expressed 30-fold or more in testis than in females, 1,897 genes were expressed 10-30 fold more in testis than in the gynogen, and 2,916 genes were expressed 5-10 fold more in the testis than in the gynogen (Table S3).

Table 3. Representative channel catfish male-biased genes involved in spermatogenesis, gonad sex determination, and testicular determination.

Feature ID	Gene name in zebrafish database	Abbreviations	Fold change	Linkage group
Contig29101	Piwi-like protein 1	PIWIL1	33.9	2
Contig22821	Transcription factor SOX-9	SOX9	7.2	3
Contig20336	IQ motif containing H	IQCH	48.8	4
Contig1684	Kinesin family member 23	KIF23	12.5	4
Contig24841	DEAD/H (Asp-Glu-Ala-Asp/His) box helicase 11	DDX11	12.2	4
Contig21605	HYDIN, axonemal central pair apparatus protein	HYDIN	31.1	4
K70:510778	Transforming growth factor beta regulator 1	TBRG1	9.3	4
Contig24257	Aquaporin 11	AQP11	7.3	4
Contig21219	Nucleoporin 93	NUP93	5.5	4
Contig17653	Cytochrome P450, family 17, subfamily A, polypeptide 1	CYP17A1	24.0	6
Contig25669	Doublesex- and mab-3-related transcription factor A2	DMRTA2	504.2	8
Contig5339	Polyadenylate-binding protein 1-like	PABPC1	614.2	11
Contig23715	Deleted in azoospermia-like	DAZL	702.4	12
Contig11277	Spermatogenesis-associated protein 22-like	SPATA22	369.3	14
Contig27135	Spermatogenesis-associated protein 17	SPATA17	50.8	16
Contig17343	Ring finger protein 17	RNF17	355.9	17
Contig22169	Anti-Mullerian hormone	AMH	11	20
Contig19377	Glutamine-dependent NAD(+) synthetase	NADSYN1	10.0	20
Contig30394	Tektin-1	TEKT1	94.4	22
Contig19309	Piwi-like protein 2	PIWIL2	362	25
Contig30210	Probable ATP-dependent RNA helicase DDX4	DDX4	146.3	25
Contig17400	Testis-specific gene 10 protein	TSGA10	265.6	26
Contig18399	Doublesex- and mab-3-related transcription factor 1 isoform 1	DMRT1	1009.5	N/A
Contig11347	Doublesex and mab-3 related transcription factor 3a	DMRT3a	50.3	N/A
Contig26052	Serine/threonine-protein kinase Nek1	NEK1	8.0	N/A
Contig9056	Protein arginine N-methyltransferase 5	PRMT5	5.6	N/A
Contig22278	RuvB-like 2	RUVBL2	8.9	N/A
Contig17332	Sperm-associated antigen 17	SPAG17	74.4	N/A
Contig6953	Sperm-associated antigen 8-like	SPAG8	42.4	N/A
Contig20873	Spermatogenesis-associated protein 4	SPATA4	37.6	N/A
Contig21931	Transforming acidic coiled-coil-containing protein 3	TACC3	5.1	N/A
K94:490875	Tudor domain containing 1	TDRD1	894.3	N/A
Contig23678	Tudor domain containing 7 isoform 1	TDRD7	404.6	N/A
Contig23675	Testis-expressed sequence 9 protein	TEX9	36.4	N/A
Contig25594	Testis-expressed sequence 14 protein	TEX14	325.8	N/A

BLASTX analysis indicated that many male-biased genes were involved in gonadogenesis, spermatogenesis, testicular determination, gametogenesis, gonad differentiation and sex determination (Table 3). Some examples of these genes include doublesex- and mab-3-related transcription factor 1 isoform 1 (Dmrt1, the sex-determining gene in medaka), Dmrta2, Dmrt3a, Amh (Amhr2, probably responsible for sex determination in fugu; amhy, a strong candidate sex-determining gene in Patagonian pejerrey), Ddx4, Ddx11, and transcription factor Sox9. Gene pathway analysis indicated that many of these genes are involved in the regulation of spermatogenesis (Figure 2).

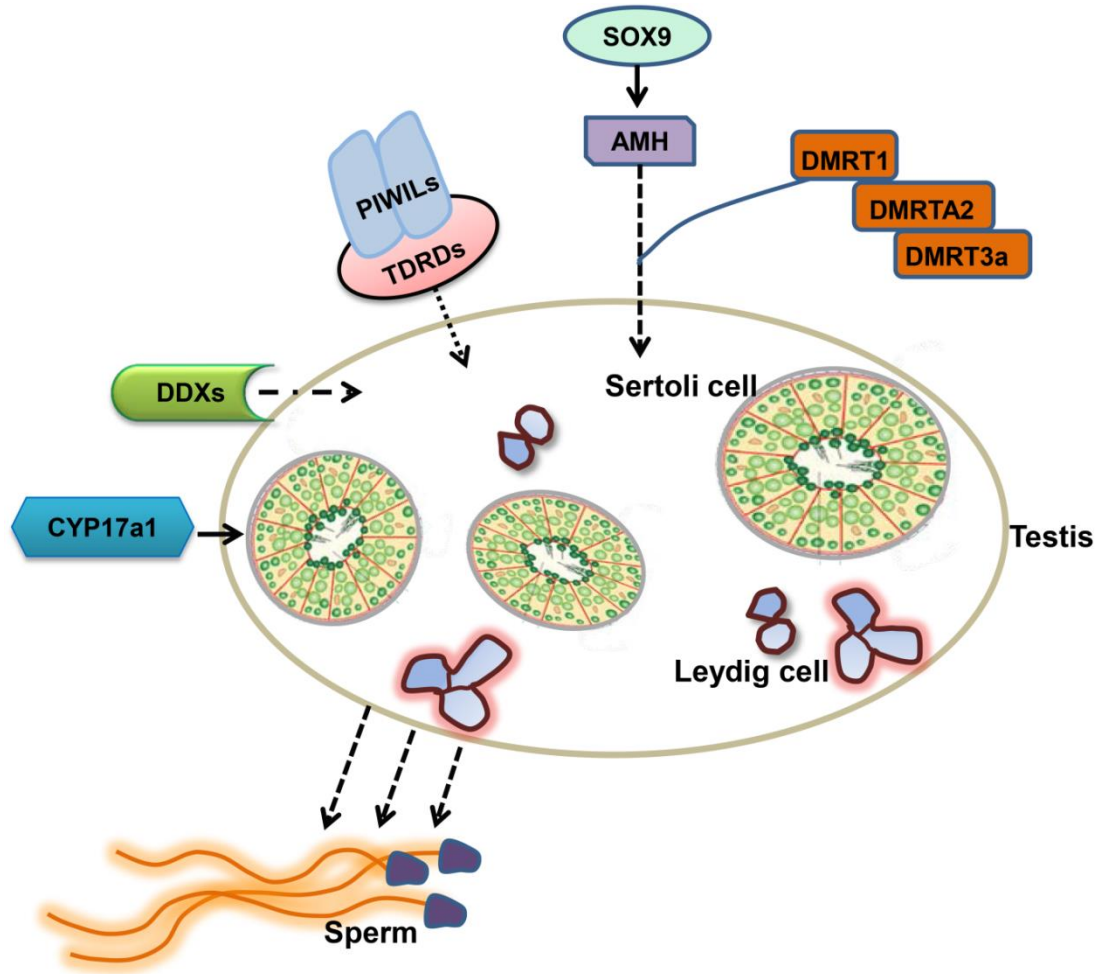


Figure 2. Putative catfish spermatogenic pathway based on RNA-Seq expression signatures in channel catfish testes.

Genomic locations of the male-biased genes were analyzed by sequence homology searches for their associated genomic draft sequence contigs. If the linkage group information is known for their associated genomic contigs, then by inference, their genomic locations are known. Analysis of genomic locations of the male-biased genes indicated that they are distributed all

over the genome among all 29 linkage groups (Figure 3). Of the 5,450 male-biased genes, 2,818 genes could be assigned to linkage groups, whereas the remaining 2,632 genes could not be assigned to linkage groups. As shown in Figure 3, the distribution of the male-biased genes is on all linkage groups. However, when the genes mostly related to spermatogenesis, gonad differentiation, and sex determination were analyzed, quite a few genes were located in linkage groups 4 (Table 3), which included the sex determination locus in catfish (Ninwichian et al., 2012b). Blast2GO program revealed a total of 31 GO terms including 7 cellular component terms, 11 molecular functions terms and 13 biological process terms. An overview of the number and the proportion of the annotated genes assigned to GO terms are shown in Figure S1.

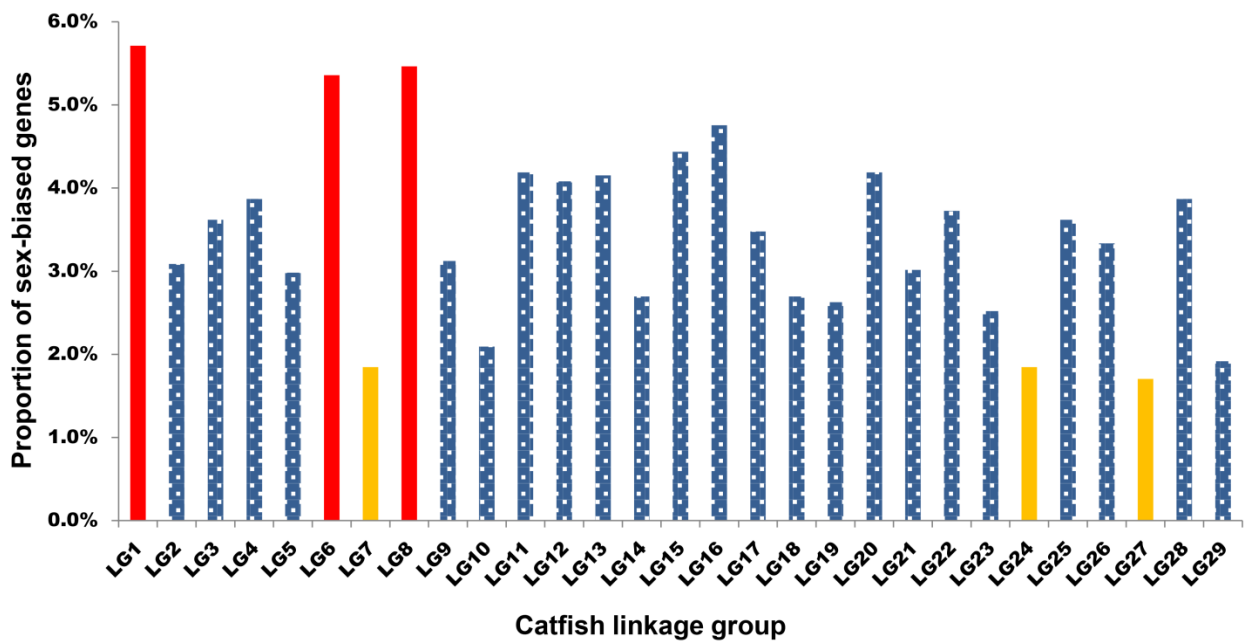


Figure 3. Linkage group distribution of the male-biased genes. Note that the proportion of male-biased genes per linkage group is similar except for chromosomes 1, 6, 8, 7, 24 and 27 (shown in red for linkage groups with high numbers of male-biased genes and in yellow for linkage groups with low numbers of male-biased genes).

Discussion

In this study, we conducted RNA-Seq analysis of the catfish testis that allowed expression profiling of genes expressed in the male-specific organ. In this process, over 25,307 unigenes were identified from the catfish testis, of which 167 unigenes were identified for the first time in catfish. This RNA-Seq analysis enhanced the transcriptome assembly in catfish, and will support the whole genome annotation of catfish.

Our long term goal is to determine the sex determination mechanisms in catfish. In addition to transcriptome profiling, we were interested in exploring the possibility of identifying male-biased genes. We therefore compared the transcriptome expression profiles in the testis to those of the gynogen female catfish. The gynogen female, a doubled haploid female, has two identical copies of each chromosome but without the Y-chromosome (Waldbieser et al., 2010). For the construction of the gynogen female transcriptome, 19 tissues were collected from a single doubled haploid female channel catfish adult, including head kidney, fin, pancreas, spleen, gill, brain, trunk kidney, adipose, liver, stomach, gall bladder, ovary, intestine, thymus, skin, eye,

swim bladder, muscle, and heart. However, no testis tissue was included because the female fish does not harbor the male-specific organ of testis. Therefore, this RNA-Seq conducted with testis tissue should enhance the transcriptome assemblies by adding on testis-specific or preferentially expressed genes.

A large number of genes were identified to express significantly higher in the male testis than in the gynogen fish. Overall, more than 5,000 had an expression ratio of greater than 5 fold in the testis than in the gynogen, however, only 637 genes were expressed much higher (>30 times) in the testis than in the gynogen fish. These putative male-biased genes can be interesting because they could be candidates contributing to gonadogenesis or testicular development and differentiation, or even sex determination. However, we realized the limitation of meta-analysis for the comparison of expression in different sexes. On the one hand, the genes important for sex differentiation are likely to be expressed in sex-specific organs such as the testis under this study. We understand that there is no direct control for such experiments because only male fish have testis. Therefore, we accepted the compromise in this initial study in order to more broadly capture the male-biased genes. In a way, the putative male-biased genes identified here can only serve as proxies of sex-biased genes. Nonetheless, a large number of genes with identities to known genes involved in gonadogenesis, spermatogenesis, testicular development and differentiation and sex determination were identified among the male-biased genes (Table 3). Interestingly, many spermatogenesis genes were believed to accumulate on the Y chromosome, suggesting their potential location in the sex-determining region (Affara and Mitchell, 2000).

Among many of the male-biased genes, a number of them are highly relevant to sex determination and differentiation, and here we discuss a few:

Dmrt genes

Of great interest in this study was the detection of the extremely high fold ratio in testis for three genes Dmrt1, Dmrta2 and Dmrt3a belonging to the Dmrt gene family. Characterized by a conserved DNA-binding motif known as the Doublesex- and Mab-3-related (DM) domain, Dmrt genes have been reported to be actively involved in sex determination and/or differentiation. These genes stimulate male-specific differentiation, but repress female-specific differentiation (Herpin and Schartl, 2011; Kopp, 2012; Wang et al., 2012). In medaka, Dmrt1 on the Y-chromosome is the master sex determination gene. The expression pattern of Dmrt genes has been studied and is clearly consistent with such related functions in southern catfish *Silurus meridionalis*, North African catfish *Clarias gariepinus*, and *Danio rerio*, suggesting the importance of this gene family during gonad development, testis sex determination, and testis differentiation (Guo et al., 2005; Liu et al., 2010; Piferrer and Guiguen, 2008; Raghuv eer and Senthilkumaran, 2009; Raghuv eer et al., 2011). In the report of Siberian sturgeon, Dmrt1 showed a significantly higher expression in testis. Similar trend of dmrt1 expression during early and advanced stages of gonad development was observed in sturgeons (Berbejillo et al., 2012), as well as in rainbow trout (Marchand et al., 2000). Dmrta2, regarded as one of the candidate genes related to sex determination and gonad differentiation, was mapped onto the linkage group containing the major sex determination factor in turbot. Even though Dmrta2 was not the

primary sex determination gene, it was thought to be involved in the sexual development of turbot (Vinas et al., 2012). The higher level of expression of *Dmrt3a*, which expressed in the same temporal and spatial pattern as *Dmrt1*, suggests its putative role in the developing gonads and sex determination (Sarre et al., 2011; Wilhelm et al., 2007). In this study, we identified *Dmrt1*, *Dmrta2*, and *Dmrt3a* among the male-biased genes. Although their function in channel catfish is unknown at present, their involvement in sex determination and differentiation needs to be further studied.

TDRDs and PIWIs

TDRD1 and TDRD7, both essential proteins for spermatogenesis, were found to be male-biased genes in this study. They were previously shown to be preferentially expressed in the murine testis (Kojima et al., 2009), and function during spermatogenesis (Hosokawa et al., 2007). Each member in the TDRD gene family performs a distinct function at different differentiation stages of spermatogenesis and TDRD7 was demonstrated to play a crucial role during early spermatid differentiation (Tanaka et al., 2011).

In this study, we also identified *PIWIL1* and *PIWIL2* as male-biased genes. *PIWIL1* and *PIWIL2* are members of the mouse Piwi family proteins (*MIWI*, *MILI*, and *MIWI2*) that play important roles in spermatogenesis through transcriptional and post-transcriptional regulation. Acting as a functional partner of Piwi family proteins, TDRD group of Tudor proteins were reported to be physiological binding partners of Piwi family proteins and coordinately work together in the regulation of spermiogenesis (Bak et al., 2011; Kojima et al., 2009). Here in our

findings, the fact that both PIWIs and TDRDs were identified among the male-biased genes is intriguing, suggesting that they may similarly work side by side in coordination of spermatogenesis. Further investigations are needed to gain understanding of the functions of these genes.

DDXs

DDX4 and DDX11 exhibited high level of expression in the catfish testis, and were both identified as male-biased genes. DDX4 and DDX11, both members of the DEAD/DEAH box family of helicases, were believed to be involved in embryogenesis, spermatogenesis, and cellular growth and division (Godbout and Squire, 1993). DDX4, as one of the 14-3-3 interactors predominantly expressed in testis, was reported to be essential for spermatogenesis due to its importance in germ cell division and maturation (Puri et al., 2008). In another study, DDX4 (the gene encoding *Drosophila* VASA homolog) was shown to play an essential role in regulating germ cell differentiation in both vertebrates and invertebrates (Saffman and Lasko, 1999). In humans, DDX4 mRNA and protein were abundantly and specifically expressed in germ cells in both sexes throughout development (Castrillon et al., 2000), suggesting a role in germ cell development. Rolland et al. performed *in situ* hybridization and detected very strong signal for DDX4 in rainbow trout testis, confirming its function in spermatogenesis regulation (Rolland et al., 2009). DDX11 was shown to be expressed ubiquitously during early embryogenesis and act as one of the necessary proteins for the later stages of spermatogenesis in the mouse testis

(Clemente et al., 2006; Cota and Garcia-Garcia, 2012). Once again, the exact function of DDX4 and DDX11 in catfish is unknown at present, but certainly worthwhile of additional study.

Sox9

Interestingly, Sox9, one of the most important genes expressed during testis determination in mammals (Jiang et al., 2012), was found to be highly expressed in the catfish testis, and as a male-biased gene. Previous studies about Sox9 in mice (Barrionuevo et al., 2006; Chaboissier et al., 2004; Sinclair et al., 1990; Vidal et al., 2001) and frogs (El Jamil et al., 2008; Takase et al., 2000) indicated its critical role in testis differentiation and development. Sox9, together with Dmrt1, is one of the key SRY targets in mammalian testis development. In mice, Sox9 was also considered to be required and sufficient for testis formation (Bagheri-Fam et al., 2010). Sox9 was also reported to be involved in vertebrate sex determination (Kent et al., 1996). In the male sex-determination pathway, Sox9 is a very early-acting gene. For instance, mutations of Sox9 have been shown to interfere with male sex determination, suggesting its role in testis determination (Koopman, 1999). Much fewer studies were conducted with Sox9 in fish. It was conserved and expressed during testicular development stage in fish, and was believed to be a candidate gene involved in testis differentiation, but not in sex determination (Berbejillo et al., 2012; Nakamoto et al., 2005). Apparently, as the male-biased gene in catfish, the functions of Sox9 in catfish testis development need to be studied.

Materials and methods

Ethics statement, experimental fish and sample collection

All procedures involving the handling and treatment of fish used during this study were approved by the Auburn University Institutional Animal Care and Use Committee (AU-IACUC) prior to initiation of experiments. The fish used for this project were 25 male channel catfish including 13 one-year-old juveniles, 8 two-year-old adults and 4 four-year-old sexual maturation adults. All fish were sexed based on external genitalia, followed by anatomical confirmation. The fish were euthanized with tricaine methanesulfonate (MS 222) at 300 mg/L (buffered with sodium bicarbonate) before sample collection. Testis tissues from the 3 different ages were placed into 5 mL RNA later™ (Ambion, Austin, TX, USA) respectively. After one day of temporary storage at 4 °C, samples immersed in the RNA later™ were transferred to a -80 °C ultra-low freezer until preparation of RNA.

RNA isolation, library construction and Illumina sequencing

Prior to RNA extraction, samples were removed from the -80 °C freezer and ground with sterilized mortar and pestle in the presence of liquid nitrogen to a fine powder. Total RNA was extracted from testis powder using the RNeasy Plus Kit (Qiagen) treated with RNase free DNase I (Qiagen) to remove genomic DNA. RNA concentration and integrity was measured on an

Agilent 2100 Bioanalyzer using a RNA Nano Bioanalysis chip. Equal amount of total RNA from catfish testis of each age was pooled together into only one sample for use in RNA-Seq.

RNA-Seq library preparation and sequencing was carried out by HudsonAlpha Genomic Services Lab (Huntsville, AL, USA) as previously described by Li et al. and Sun et al. (Li et al., 2012; Sun et al., 2012). cDNA library was prepared with ~4 µg of starting total RNA following the protocols of the Illumina TruSeq RNA Sample Preparation Kit (Illumina). The library was amplified with 15 cycles of PCR. The final library had an average fragment size of ~270 bp and final yields of ~400 ng. After KAPA quantitation and dilution, the library was sequenced with one lane on an Illumina HiSeq 2000 instrument with 100 bp paired end (PE) reads. Raw read data of channel catfish testis RNA-Seq are archived at the NCBI Sequence Read Archive (SRA) under Accession SRP018265 and now released.

De novo assembly of sequencing reads

Before *de novo* assembly, raw sequencing reads were trimmed by removing adaptor sequences, ambiguous nucleotides ('N' at the end of reads), low quality sequences (quality score less than 20), and short read length sequences (length below 30 bp) with CLC Genomics Workbench (version 4.8; CLC bio, Aarhus, Denmark) as previously described (Li et al., 2012; Sun et al., 2012). The assembly was performed using the *de Bruijn* graph approach with ABySS (version 1.3.4) (Simpson et al., 2009) and Trans-ABySS version 1.3.2 to obtain accurate and reliable consensus contigs as a reference assembly (Robertson et al., 2010). Briefly, continuous multiple k-mers ranging from 50 to 96 were used in ABySS, and then all 47 assemblies from

ABySS were merged into one assembly to generate the transcriptome assembly using Trans-ABySS. Afterwards, CAP3 (Huang and Madan, 1999) was utilized in order to minimize redundancy and the resulting contigs that were > 200 bp were regarded as final non-redundant transcripts. The threshold was set as 100 bp for the minimal overlap length and 99% for the identity in CAP3.

Transcriptome annotation and ontology

The assembled contigs including both unique transcripts and singletons were used as queries for BLAST searches against the zebrafish RefSeq protein database, UniProtKB/SwissProt database and non-redundant database, respectively. Searches were conducted using the BLASTX program, with an E-value cut-off of $1e-10$ and matching to the top hits. Gene ontology (GO) annotation analysis was performed using the zebrafish BLAST results in Blast2GO version 2.5.1, which is an automated tool for the assignment of gene ontology terms. The zebrafish BLAST result was imported to Blast2GO. The final annotation file was produced after GO-mapping, GO term assignment, annotation augmentation and generic GO-Slim process. The annotation result was categorized with respect to Biological Process, Molecular Function, and Cellular Component at level 2.

Identification of putative novel transcripts

In order to identify putatively novel transcripts from the currently assembled testis transcriptome, the assembled contigs were used as queries to search against the existing catfish transcriptome databases (Li et al., 2012; Liu et al., 2012b; Sun et al., 2012), using blast with E-value cutoff of $1e-5$. Those transcripts without blast hits after *in silico* subtraction were identified as putative newly identified transcripts.

Identification of the male-biased genes

To identify the sex-biased transcripts, high quality sequence reads from the male catfish testis RNA-Seq reads (current work) and the female doubled haploid catfish RNA-Seq reads (Liu et al., 2012b) were mapped onto the testis transcriptome using CLC Genomics Workbench respectively. At least 95% of the read length was required to align to the reference and a maximum of two mismatches were allowed during mapping. The unique gene reads number for each transcript was determined, and then normalized to RPKM (Reads Per Kilobase per Million). The proportions-based Kal's test for differences was used to identify the differentially expressed genes between testis and the gynogen female with p-value < 0.05 . The fold changes were calculated after quantile normalization of the RPKM values. Transcripts with absolute fold change values of larger than 5 and total read number larger than 10 were included in analysis as the sex-biased genes.

Gene ontology (GO) annotation analysis was performed on the preferentially expressed genes in testis using Blast2GO version 2.5.1. The final annotation file was produced after GO-mapping, GO term assignment, annotation augmentation and generic GO-Slim process. The

annotation result was categorized with respect to Biological Process, Molecular Function, and Cellular Component at Level 2.

In order to determine the genome distribution, the male-biased genes were mapped to the catfish linkage groups by taking advantage of the existing genomic resources, especially the integrated genetic linkage and physical map (Ninwichian et al., 2012a).

Acknowledgements

We would like to thank Dr. Eric Peatman and Dr. Huseyin Kucuktas for their assistance in providing some of the fish and sampling of tissues. We appreciate the excellent sequencing services of HudsonAlpha Genomic Services Lab (Huntsville, AL, USA). We are grateful to Alabama Supercomputer Center for providing the computing capacity for the bioinformatics analysis of this study.

**CHAPTER IV. ASSEMBLY AND COMPARATIVE ANALYSIS OF
THE X AND Y CHROMOSOMES IN CHANNEL CATFISH, AND IDENTIFICATION
OF THE PUTATIVE CATFISH SEX-LINKED REGION**

Abstract

Catfish have a XY male/XX female sex chromosome system. The exact mechanism of sex determination in catfish is unknown at present. This could be not only attributed to the extraordinary plasticity and diversity of the teleost sex determination mechanism, but also the incompleteness of assembly and annotation of the catfish sex chromosome itself.

To begin to fill the gaps, here we utilized multiple approaches to get the best assembly and scaffolding of X and Y chromosome in catfish and conducted *in silico* comparative analysis between these two chromosomes. Sequencing of the super-male (YY) sample, the regular male (XY) sample, and BAC clones on the sex chromosome using the Illumina HiSeq 2000 platform followed by assembly by ABySS and scaffolding by SSPACE generated 100,173 scaffolds with an average length of 9,060 bp and N50 of 91,893 bp for the male whole genome. The Y chromosome is approximately 77 Mb in size covering by 1,410 scaffolds. On the other hand, female sex chromosome contained 1,735 scaffolds based on our latest version of female whole genome. Annotation using BLASTX allowed the identification of 497 unigene hits against the NCBI *nr* database in both the Y chromosome and the X chromosome.

Further comparative analyses were performed with the rapidly genome aligning software MUMmer. However, no Y-specific sequence was identified. The present work demonstrates that there is no gene deletion between X and Y chromosome in channel catfish. In addition, based on the position of the sex-linked marker, a 203 kbp region on Linkage Group 4 covering 16 genome contigs was identified and regarded as the putative catfish sex-linked region. This region

encompasses 10 known genes, with one strong candidate “ZBTB38” that is involved in sexual orientation in male *Drosophila*. Follow-up experiments will focus on the gene expression signatures of the genes in the putative catfish sex-linked region during early embryonic stages with the goal to pinpoint the expression candidates of sex-determining genes.

Key words:

Catfish; sex chromosome; assembly and scaffolding; comparative analysis; putative SD region

Introduction

Catfish, one of the lower teleosts, is the dominant aquaculture species in the United States taking up more than 60% of U.S. aquaculture production (Hanson and Sites, 2013). Channel catfish, as a good research model, are native to the Mississippi River drainage and their range has expanded to most regions of North America. One of the most fundamental features in the life of a channel catfish is the sex. It usually exhibits extraordinary plasticity and diversity among different fish species. In addition, sex also serves as an important trait that associates with the growth since male catfish grow faster than females in the wild (Kelly 1996). All the above make it worthwhile to study and illustrate the mechanism of catfish sex determination.

Channel catfish normally possess a relatively simple genetic sex determination system (XX/XY male heterogametic system) (Davis et al., 1990; Tiersch et al., 1990; Tiersch et al., 1992), but high temperature extremes applied during the critical period for the sex differentiation could cause female-skewed sex ratios. This suggested that environmental factors, such as temperature, can influence the sex determination of catfish as well (Patino et al., 1996). Previous molecular genetics analysis was conducted with genes such as *Sry* (mammalian sex-determining gene) and its closely linked gene *Zfy* in channel catfish. The results showed that even though these genes existed in the catfish genome, they were not seen to express in a sex-specific pattern (Tiersch et al., 1992). An isozyme locus GPI-B that was very close to the centromere was found on the sex chromosomes in channel catfish. This locus was located approximately 16 map units away from the sex-determination locus and on the other side of the centromere (Liu et al. 1996).

Seven microsatellite loci closely linked with the sex were identified when constructing the microsatellite-based linkage map (Waldbieser et al., 2001). However, these loci were restricted to a specific catfish family and can't be widely applied to other families or strains. Recently, Ninwichian et al. reported the identification of one Y-linked specific microsatellite marker, which displayed 100% sex-typing accuracy from four different strains of channel catfish. This method has already been put into use as a rapid and efficient way to identify the sex of channel catfish in the lab. PCR amplifications of the sex-linked marker locus produced both a male-specific (Y-linked) product and an autosomal product seen in both males and females, which may lead us to consider whether or not there is a Y-specific region (Ninwichian et al., 2012b). Since the sex-related marker was identified, the chromosome that contains the sex-linked marker was designated as the sex chromosome. Then how far the marker is away from the sex-determining gene is still unknown.

The sex-determining gene identification in fish species is being accelerated through the application of both molecular/genetic manipulations and bioinformatics approaches (Piferrer et al. 2012; (Kikuchi and Hamaguchi, 2013). Sex determination mechanisms have been so far elucidated in five fish species, including medaka (*Oryzias latipes*) (Matsuda et al., 2002; Nanda et al., 2002), rainbow trout (Yano et al., 2012), fugu (Kamiya et al., 2012), Patagonian pejerrey (*Odontesthes hatcheri*) (Hattori et al., 2012) and another species of medaka (*O. luzonensis*) (Myosho et al., 2012). The first well studied species among fish is medaka (*O. latipes*) with respect to sex determination. *Dmy* or *dmrt1bY* gene was reported to be the responsible gene for the sex determination in medaka. It is a duplicated gene of *dmrt1* on the Y-chromosome

(Matsuda et al., 2002; Nanda et al., 2002). However, the SD gene *dmy* is not conserved and differs between even close species. $Gsdf^Y$ (gonadal soma derived growth factor on the Y chromosome) has replaced the role of *dmy* as the master sex-determining gene in *O. luzonensis*. Using positional cloning technique, this gene was demonstrated to be responsible for male-specific high expression during sex differentiation. Furthermore, the presence of $Gsdf^Y$ gene fragment converted XX individuals into fertile XX males. All evidence elucidated the role of $Gsdf^Y$ on the SD cascade in the Oryzias fishes. In rainbow trout, *sdY* gene, a truncated immune-related gene interferon regulatory factor 9 like (*Irf9*) gene residing on the Y-chromosome, was found to be the sex control gene. Divergent from the *Irf9* protein, *sdY* may obtain a new function in the gonadal sex determination (Yano et al., 2012). Both family-based genetic mapping and association mapping in fugu allowed the identification of the missense sex-determining single nucleotide polymorphism, which located within the anti-Mullerian hormone receptor type II (*Amhr2*) gene (Kamiya et al., 2012). It is noteworthy that the SD locus in fugu resides in a recombining region, which is contrary to the situation in medaka *dmy* (Kondo et al., 2006) and therian *Sry* (Graves, 2006). With Patagonian pejerrey, *Amhy* gene, which is a male-specific, functional duplicated copy of the anti-Mullerian hormone gene (*amh*) was suggested to be involved in the sex determination in this species (Hattori et al., 2012). In addition to these where the sex determination gene has been identified, the SD region has been found in several other fish species, such as stickleback (*Gasterosteus aculeatus*) (Peichel et al., 2004), turbot (Martinez et al., 2009), Nile tilapia (*Oreochromis niloticus*) (Eshel et al., 2012; Lee et al., 2011), Japanese flounder (*Paralichthys olivaceus*) (Castano-Sanchez et al., 2010), half-smooth tongue

sole (*Cynoglossus semilaevis*) (Shao et al., 2010) and Pacific halibut (*Hippoglossus stenolepis*) (Galindo et al., 2011).

Learning from the success in identification of the sex-determining gene in fish species, here, we come up with two major approaches (using XY system as an example): 1) Identification of Y-specific sequences; 2) Identification of male-specific transcripts. In both cases, the sex-determining gene must be validated by functional analysis of the Y-specific sequences and male specific transcripts for their necessity and sufficiency for sex determination. Unlike the obviously morphological distinct sex chromosomes (X and Y) in mammals, X chromosome and Y chromosome in channel catfish are quite similar (Tiersch et al., 1990). However, there may exist the difference in the DNA content (DNA sequences) from the X and Y chromosome in channel catfish. The issue is how we can identify this region. In the present study, we took the advantage our catfish physical map (Quiniou et al., 2007; Xu et al., 2007), linkage map (Kucuktas et al., 2009; Ninwichian et al., 2012a; Waldbieser et al., 2001), and genome resources to with an aim to assemble and scaffold the X and Y chromosome (two differentiated channel catfish heteromorphic sex chromosomes), followed by comparing these two. Our goal is to answer the question if or not there is a Y-specific sequence in channel catfish.

Materials and methods

Sample preparation and sequencing of the regular male (XY)

All procedures involving the handling and treatment of fish used during this study were approved by the Auburn University Institutional Animal Care and Use Committee (AU-IACUC) prior to initiation. A two-year-old male channel catfish, weighing 1.12 kg was used for DNA extraction. Genomic DNA was extracted from its blood using a Qiagen DNeasy Blood and Tissue kit (cat. 69504). The concentration of genomic DNA was measured using an Agilent 2100 Bioanalyzer. DNA integrity was checked through an easy and fastest way which is to use 0.8% agarose gel, run it at 80v for 45 minutes.

DNA from the male channel catfish was directly sequenced using the Genome Sequencing Service provided by Iowa State University. Following the library preparation containing 300-bp inserts, the sequencing reactions were performed using Illumina HiSeq 2000 instrument. Sequencing was performed with the paired-end strategy of 100-bp reads, representing the whole genome.

Sample preparation and sequencing of the BAC clones on the putative sex chromosome

Since the Y-linked specific microsatellite marker was identified from Linkage Group 4 in channel catfish, we decided to capture the whole chromosome with the sex-linked marker. Based on catfish linkage map and physical map, catfish BAC clones that covered the minimum tiling path in each physical contig on Linkage Group 4 were picked up from the CHORI-212 BAC library (Xu et al., 2007).

Each well on the 96-well culture block was filled with ~1 ml of 2×YT medium with 12.5 µg/ml chloramphenicol. BAC clones were then transferred from the 384-well plate to the 96-

well culture block. The whole plate was then shaken up with the speed of 300 rpm at 37°C overnight to make the bacteria grow. The block was centrifuged at 2000× *g* for 10 min in an Eppendorf 5804R bench top centrifuge to collect bacteria. The culture supernatant was decanted and the block was inverted and tapped gently on paper towels to remove remaining liquid. BAC DNA was isolated using the Perfectprep™ BAC 96 kit (Eppendorf North America, Westbury, NY) according to the manufacturer's specifications (Jiang et al., 2011). The concentration of BAC DNA was measured using an Agilent 2100 Bioanalyzer and the DNA integrity was checked by electrophoresis on agarose gel. High quality pooled BAC DNA was sent to Genome Sequencing Service provided by Iowa State University for sequencing using Illumina HiSeq 2000.

Sample preparation and sequencing of the super male (YY)

The sex-reversed YY super male was provided by Dr. Waldbieser. The blood sample was flash frozen in liquid nitrogen and shipped on dry ice then stored at -80 °C until DNA extraction. Genomic DNA was extracted using a Qiagen DNeasy Blood and Tissue kit (cat. 69504). The concentration of genomic DNA was measured using an Agilent 2100 Bioanalyzer. DNA integrity was checked through an easy and fastest way which is to use 0.8% agarose gel, run it at 80v for 45 minutes as previously described.

DNA-seq library preparation and sequencing was carried out by HudsonAlpha Genomic Services Lab (Huntsville, AL, USA). The DNA library that had an average fragment size of

~300 bp was prepared and amplified with PCR. Sequencing with 100-bp paired end (PE) reads was performed on an Illumina HiSeq 2000 instrument.

De novo assembly of three sets of data

Low quality reads (reads with quality scores less than 20), BAC vector sequences and reads shorter than 20 bp were discarded using CLC Genomics. As the dominant software for genome sequencing, the *de bruijn* graph based assembler ABySS was utilized for assembly. Multiple-k-mer assemblies (k-mer sizes from 51 to 96) were produced. The erode-strand was set E=0 and scaffold option off. The minimum number of pairs needed to consider joining two contigs was set to 10. CAP3 (Huang 1999) were used to remove assembly redundancy with the minimal overlap length and percent identity in CAP3 to 100 bp and 99%.

In order to construct an optimized assembly backbone for genome scaffolding, the assembly results of the regular male (XY) genome, sex chromosome BAC clone sequences and the super male (YY) genome were compared. The best assembly was evaluated according to parameters including the number of contigs generated, N50 size, average contig size, and maximum contig length. The resulting best assembly was used in subsequent scaffolding.

Scaffolding of the male genome

The scaffolding was performed by five schemes in order to obtain the best scaffolding results. Taking advantages of all available paired-end sequencing reads (with different insert lengths of the paired-end libraries: 300bp, 3kb, 8kb and 36kb respectively) and the optimized male

assembly result, we built up the male genome scaffolding using the publicly available programs SSPACE (version basic v1.0) and SOAPdenovo (version 2.04).

ABYSS assembly was utilized as a backbone for SSPACE programming. The clean paired-end reads were used as input to integrate the pre-assembled contigs and a stand-alone scaffolding was performed (Boetzer et al., 2011). The novel short-read assembly program SOAPdenovo was employed to construct the draft scaffolding of male genome as well (Luo et al., 2012). SOAPdenovo was well known to assemble the Illumina short reads and used as a reference to examine and supplement the scaffold result obtained by SSPACE. The detailed schemes for male genome scaffolding are presented as following:

- 1) Super male (YY) ABySS *de novo* assembly was used as a backbone and a stand-alone scaffolding was performed of pre-assembled contigs by SSPACE with the gynogen 3K, 8K and 36K high quality paired-end reads. The parameters were set as $k=3$, $a=0.7$, $x=0$, $r=0.8$, $g=3$. Here, 'k' stands for minimal links and 'a' is the maximum link ratio. These two parameters played vital roles in the scaffolding. The k option specifies the minimum number of links (read pairs) a valid contig pair must have to be considered. The 'a' option specifies the maximum ratio between the best two contig pairs for a given contig being extended. Where the 'x' is contig extension option which indicates whether to do extension or not. If 'x' is set to 1, contigs are tried to be extended using the unmapped sequences. If set to 0, no extension is performed. The 'r' stands for minimal base ratio which is used to accept an overhang consensus base. Higher 'r' value lead to more accurate contig extension. The 'g' is maximum gaps allowed for Bowtie, and this parameter is used both at mapping during extension and mapping during scaffolding. This parameter is

suggested to be increased when large reads are used, e.g. Roche 454 data or Illumina 100bp (Boetzer et al., 2011).

2) Super male (YY) ABySS *de novo* assembly served as a backbone for scaffolding. The paired-end reads of the super male (YY) was utilized as the input in SSPACE. The parameters are set as $k=3$, $a=0.7$, $x=0$, $r=0.8$, $g=3$.

3) The high quality pair-end reads of super male (YY) as well as the gynogen 3K, 8K and 36K PE reads were used for *de novo* assembly by SOAPdenovo and further scaffolding. In addition, Moleculo long reads from Blue Catfish genome (unpublished data) with an average length of 4,600bp were also used in the scaffolding step due to the capability of linking proximal contigs together. The parameters were set as '-K 63 -R -F', where K specified K-mer size 63 was selected based on the previous sequence assembly information from ABySS *de novo* assembly; R indicated the option to resolve repeats by reads, and F stands for the option to fill gaps in scaffold based on the super male (YY) high quality pair-end sequences (Luo *et al.* 2012).

4) The high quality pair-end reads of super male (YY) were used for both *de novo* assembly and further scaffolding steps through SOAPdenovo. The parameters are set as '-K 63 -R -F'.

In order to make it clear, here we explain the schematic of catfish Y chromosome assembly and scaffolding in Fig. 1.

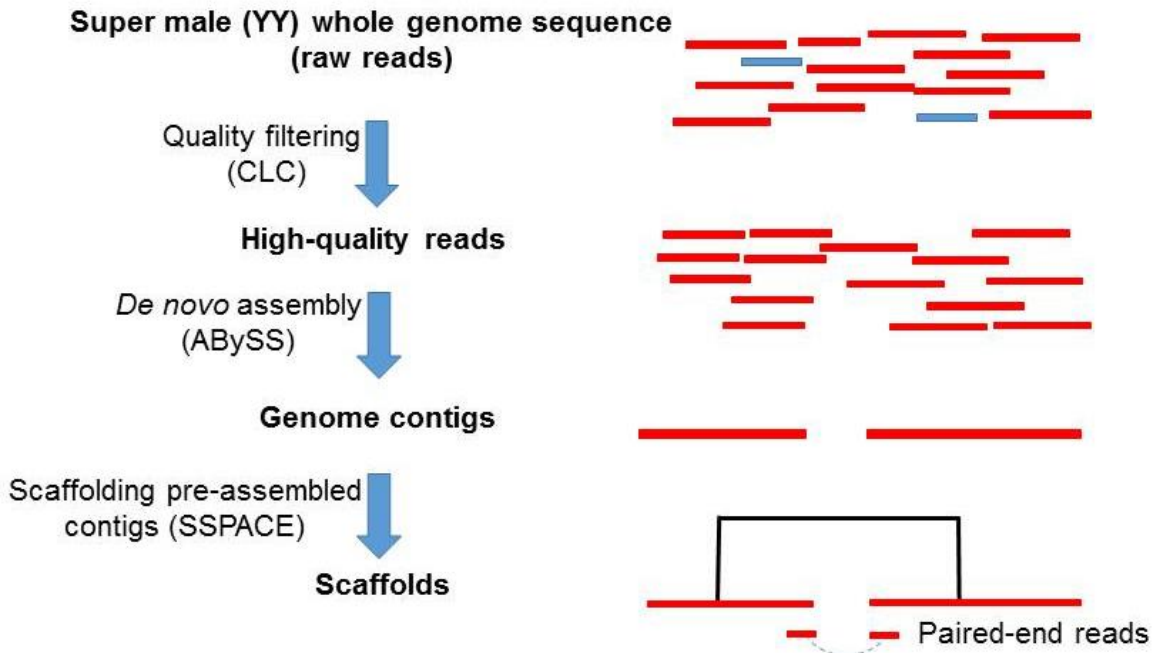


Fig. 1. A schematic overview of catfish Y chromosome assembly and scaffolding

Raw reads of super male (YY) whole genome were trimmed using CLC and assembled with the optimized k-mer size using ABySS assembler. Using paired-end sequencing data with different insert sizes of the Illumina paired-end libraries, the final genome contigs were used as a backbone for scaffolding (assess the order, distance and orientation of contigs and combine them into scaffolds).

Screening the scaffolds on the sex chromosome out of the whole genome scaffolding

Queries for capturing the scaffolds on the sex chromosome include three parts: physical map contig-specific sequences on Linkage Group 4 (unpublished work by Jiang et al., 2013), sex chromosome BAC clone sequence assembly, and BAC-end sequences on Linkage Group 4. These sequences were used as queries to search against the draft catfish genome scaffolds by

BLASTN program, with an E-value cutoff of $1e-20$. The query sequences with multiple hits of genome contigs were considered non-specific and discarded. Only the query sequences with a single hits and identity value greater than 98% were considered as real sequences on the sex chromosome. The corresponding genome contig were retrieved for further alignment analysis.

Comparative analysis of X and Y chromosome sequences

In order to identify the Y-specific fragments, a comparative analysis was performed with the rapidly genome aligning software MUMmer. Because of the higher coverage of sequencing of the gynogen (XX) female channel catfish, average lengths of the scaffolds on the X chromosome was much larger than that on the Y chromosome. Here, X chromosome was used as a reference, and the Y chromosome was used as a query and streamed against the X chromosome.

Genome-wide comparison between super male catfish genome and female genome

With the aim to identify the Y-specific genes from the whole genome, a comparative analysis was performed between male catfish genome and the draft female genome by BLASTN program. Super male (YY) whole genome assembly contigs and singletons were used as queries, and the latest version of the draft female genome scaffolds and degenerated sequences served as the subjects. E-value cutoff was set to $1e-10$, and the identity cutoff was set more than 96%. MUMmer was employed to perform the genome-wide alignment as well.

Identification of the putative catfish sex-linked region

The position of the putative catfish sex-determining region was estimated by assuming that the sex-linked marker is located on the sex chromosome and the sex-determining gene is somewhere located near the sex-linked marker (Ninwichian et al., 2012b). Taking advantage of second generation of catfish genetic linkage map (Ninwichian et al., 2012a), the microsatellite markers near the sex-linked one were identified. The genotype of F2 hybrid family fish, F3 hybrid family fish, and pure channel catfish were reanalyzed using previous data. Based on the recombination frequency, the putative catfish sex-linked region was identified as indicated.

Annotation of the genes located in the sex-linked region

The corresponding genome sequences were retrieved using the BAC-end marker sequences. The assembly contigs were used as queries against the NCBI non-redundant (nr) protein database using the BLASTX program. The cutoff E-value was set at $1e-5$ and only the top gene id and name were initially assigned to each contig. The cutoff identity value was set greater than 98%.

Results and Discussion

Sequencing of short reads from regular male (XY) catfish

Illumina-based high throughput sequencing was carried out on the regular male (XY) blood sample. A total of 100.6 million 100 bp PE reads were generated on an Illumina HiSeq 2000 instrument in a single lane. After removing ambiguous nucleotides, low-quality sequences

(quality scores <20) and sequences less than 30 bp, 92.3% (92.9 million) of the short reads were preserved (Table 1).

Sequencing of short reads from BAC clones on the putative sex chromosome

A total of 415 BAC clones were picked up from Linkage Group 4. One lane of Illumina-based high throughput sequencing was carried out on the pooled BAC clone DNA sample. Approximately 110.4 million 100 bp PE reads were generated on an Illumina HiSeq 2000 instrument in a single lane. After removing ambiguous nucleotides, low-quality sequences (quality scores <20) and sequences less than 30 bp, 92% (101.5 million) of the short reads were preserved (Table 1).

Sequencing of short reads from super male (YY) catfish

Illumina-based high throughput sequencing was carried out on the super male (YY) blood sample. A total of 369.3 million 100 bp PE reads were generated on an Illumina HiSeq 2000 instrument in a single lane. After removing ambiguous nucleotides, low-quality sequences (quality scores <20) and sequences less than 30 bp, 96.1% (355 million) of the short reads were preserved (Table 1).

Table 1. Summary of Illumina sequencing data and metrics of the ABySS *de novo* assembly results of male catfish resources.

	Genome sequencing of regular male (XY)	Sequencing BAC clones on Y chromosome	Genome sequencing of super male (YY)
Number of reads	100,631,412	110,412,042	369,314,362
Avg. read length (bp)	103	101	100
Number of reads after trimming	92,888,092	101,539,953	355,074,883
Percentage retained	92.3%	91.96%	96.14%
Avg. read length after trimming (bp)	81	71	91
Number of contigs	607,376	42,173	292,833
Avg. length (bp)	273	380	5,109
N50 (bp)	266	400	2,539

Best assembly selection with a potential to be the scaffold backbone

In a comparison of the assemblies among the three different sets of sequencing data using ABySS (Table 1), assembly result of super male (YY) genome sequences was considered to be the best assembly that could serve as a backbone for the scaffolding. YY assembly had a relatively small number of contigs, the longest average contig size, and the longest N50. Therefore, the ABySS assembly of super male (YY) genome sequences which contained 292,833 contigs with average length of 5,109 bp was selected for subsequent male genome scaffolding.

Best scaffolding selection of the male genome

In a comparison of the metrics of the scaffolding results under four schemes of male genome (Table 2), scaffolding results with various insert size (3K, 8K and 36K high quality paired-end reads) by SSPACE provided the best scaffolding results. A total of 100,173 scaffolds with N50 size of 91,893 bp and average length of 9,060 bp were generated. Considering the indexes

including the number of scaffolds generated, N50 size, and the average length of the scaffolds, the scaffold version with four different insert sizes by was selected for subsequent Y-chromosome sequence identification.

Table 2. Summary of the metrics of the male genome scaffolding results from five scaffolding schemes

Data sources	Software	Number of scaffolds	N50 (bp)	Avg. length (bp)
YY whole genome sequence	SSPACE	100,173	91,893	9,060
XX whole genome sequence (3k insert)				
XX whole genome sequence (8k insert)				
XX whole genome sequence (36k insert)				
YY whole genome sequence	SSPACE	173,117	9,317	4,305
YY whole genome sequence	SOAPdenovo	1,603,651	8,367	832
XX whole genome sequence (3k insert)				
XX whole genome sequence (8k insert)				
XX whole genome sequence (36k insert)				
Blue XX whole genome sequence				
YY whole genome sequence	SOAPdenovo	1,263,704	4,818	724

Comparison of the sex chromosome of male and female catfish

A total of 1,735 scaffolds were identified to have significant identity with the queries including physical map contig-specific sequences on Linkage Group 4, sex chromosome BAC clone sequence assembly, and BAC-end sequences on Linkage Group 4 from the female genome scaffolding. While, 1,410 scaffolds from the male genome scaffolds have significant BLASTN hits against the queries. The alignment result showed that all scaffolds from the Y chromosome

can be mapping to the X chromosome, which indicated that there is likely no Y-specific fragment in channel catfish (Fig. 2, Table 3). In another way of saying, probably there is no extra gene in the catfish SD region that contains a distinct male-specific segment. Therefore, the X chromosome and Y chromosome are possibly undifferentiated. The sex of catfish could probably be controlled by a very tiny difference in the genomic sequence.

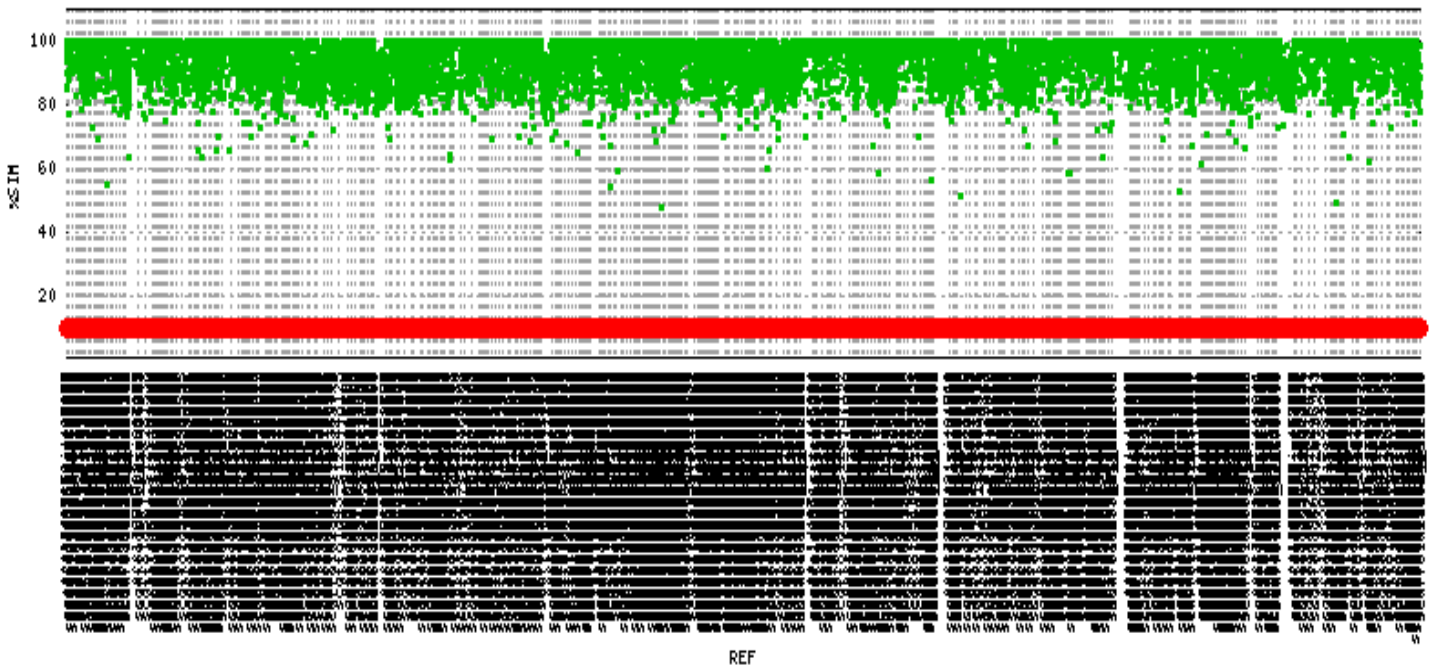


Fig. 2. Output of MUMmer sequence alignment package to identify Y-specific regions in the large sequence sets on the Y chromosome. REF stands for the reference genome sequences when performing the alignment (Channel catfish X chromosome sequence); % SIM indicates percent similarity of the alignment.

Table 3. Summary of X chromosome and Y chromosome of channel catfish

	X chromosome	Y chromosome
Number of scaffolds	1,735	1,410
N50	165,278	194,422
Avg. scaffold length	30,478	46,224
Total bases	64,828,653	77,353,140
Aligned sequences	1,712 (98.67%)	1,410 (100%)
Unaligned sequences	23 (1.33%)	0 (0 %)
Avg. identity	96.14	96.14

Comparison of super male (YY) genome and the draft female catfish genome

A total of 292,833 super male (YY) genome contigs and singletons were aligned to the draft female genome scaffolds. Approximately 50k sequences couldn't be mapped onto the female genome. Among them, around 30k sequences showed significant BLASTN hits against the degenerated sequences in the female genome. BLASTX program was employed to annotate the rest 19,699 putative Y-specific. However, none of them seem to be a real gene based on the NCBI non-redundant (nr) protein database. Most of these putative Y-specific fragments are very short, or contain a large portion of repetitive elements such as microsatellites. Another possibility is that they were generated by the assembler with a certain error rate. Therefore, no Y-specific genes were identified from the current version of super male genome and the draft female genome.

Analysis of the putative catfish sex-linked region

The result of chromosome-wide comparison between X chromosome and Y chromosome led us to focus on the putative sex-linked region more. Deep investigation of the putative catfish sex-linked region revealed the sex-linked marker-centered area that takes up approximately 203 kbp genomic sequences. Ten microsatellite markers from BAC-end sequences were included within this region (Fig. 3). A total of 16 genome contigs were captured from the draft genome scaffolds. Ten known genes that were previously annotated in other organisms were identified by BLASTX program (Table 4). Among them, “ZBTB38” was thought to be a strong candidate for the sex determination because it was reported that a BTB domain containing zinc finger protein was involved in the male *Drosophila* sex orientation (Ito et al., 1996).

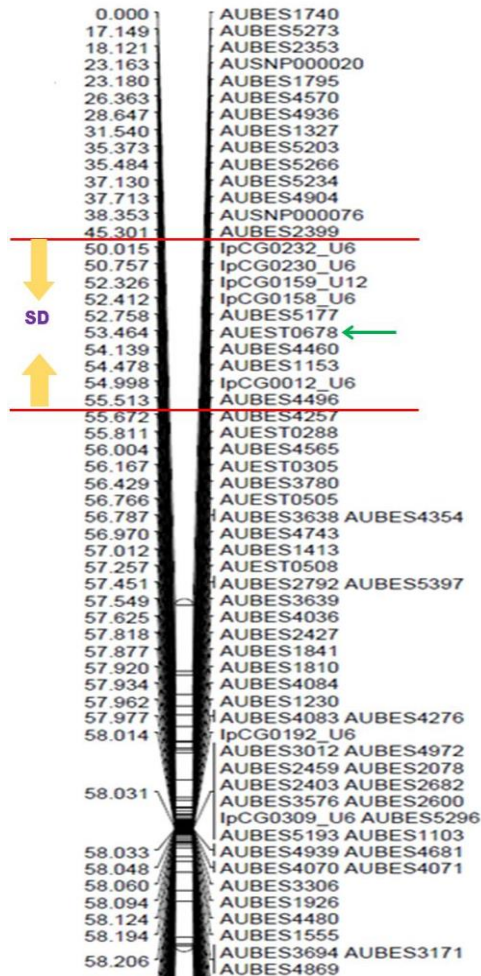


Fig. 3. Partial linkage map of catfish Linkage Group 4 with the putative sex-linked region. The putative sex-linked region (SD) was defined the region between the interval of two markers: IpCG0232_U6 and AUBES4496. The green arrow points out the sex-linked marker AUEST0678 identified in previous work (Ninwichian et al., 2012b).

Table 4. Gene list in the putative sex-linked region of channel catfish

Accession No.	Description	Organism	E_value
ABV31710.1	Transposase	<i>Salmo salar</i>	8.00E-65
XP_002942163.1	PREDICTED: zinc finger and SCAN domain-containing protein 21-like	<i>Oreochromis niloticus</i>	6.00E-14
XP_003200428.1	PREDICTED: retrotransposable element Tf2 155 kda protein type 1-like	<i>Esox lucius</i>	4.00E-09
AAD19348.1	Reverse transcriptase-like protein	<i>Takifugu rubripes</i>	6.00E-93
NP_001121805.1	Probable ATP-dependent RNA helicase DHX34	<i>Danio rerio</i>	3.00E-83
ACO51862.1	Transposable element Tc1 transposase	<i>Rana catesbeiana</i>	3.00E-20
XP_002195287.1	PREDICTED: zinc finger and BTB domain containing 38	<i>Taeniopygia</i>	E-155
F1QH17.2	Protein-methionine sulfoxide oxidase MICAL-3	<i>Danio rerio</i>	0
ADO28302.1	Purpurin	<i>Ictalurus furcatus</i>	3.00E-17
ABB87033.1	Inwardly-rectifying potassium channel	<i>Takifugu rubripes</i>	E-141

CHAPTER V. SUMMARY AND FUTURE DIRECTIONS

Summary

Sex is considered to be a luxury because it is time and energy consuming. However, it is an essential process to generate genetic diversity through recombination. Fishes exhibit particular sexual diversity and plasticity among vertebrates. Except a few fish species, the mechanism of sex determination in fish, especially the genetic-orientated sex determination mechanism, is still largely unknown.

This study is our very first analysis towards the identification of sex-determining genes in catfish. Our idea was to achieve the ultimate goal through two major approaches: 1) Identification of Y-specific gene sequences; 2) Identification of male-specific transcripts. Starting from the male-specific organ testis, male-biased genes identified are expected to provide a candidate pool for the expression candidates involved in sex determination. A comprehensive testis transcriptome was built consisting of 193,462 contigs (N50 length: 806 bp). Among these contigs, 67,923 contigs had hits to a set of 25,307 unigenes, including 335 genes newly identified in catfish. A meta-analysis of expressed genes in the testes and in the gynogen allowed the identification of 5,450 genes that are preferentially expressed in the testes, providing a pool of putative male-biased genes. Gene ontology and annotation analysis suggested that many of these male-biased genes were involved in gonadogenesis, spermatogenesis, testicular determination, gametogenesis, gonad differentiation, and possibly sex determination. This study would lay the basis for sequential follow-up analysis of genes involved in sex determination and differentiation in catfish.

In order to identify the Y-specific gene sequences, super male (YY) genome was assembled and scaffolded using limited sources of data with relatively low sequencing coverage. A total of 100,173 scaffolds were generated with an average length of 9,060 bp and N50 of 91,893 bp. Using the *in silico* comparative analysis, sequences from the male were mapped to the draft female genome with both BLASTN program and the rapidly genome aligning software MUMmer. However, no Y-specific genes were identified based on our current version of male genome sequences. The present work indicates that there is no gene deletion between X and Y chromosome in channel catfish. In addition, based on the position of the sex-linked marker, a 203 kbp region on Linkage Group 4 covering 16 genome contigs was identified and regarded as the putative catfish sex-linked region. This region encompasses 10 known genes, with one strong candidate “ZBTB38” that is involved in sexual orientation in male *Drosophila*. Follow-up experiments will focus on the gene expression signatures of the genes in the putative catfish sex-linked region during early embryonic stages with the goal to pinpoint the expression candidates of sex-determining genes.

According to the observations in this study, there might be no extra gene in the male catfish that control its sex determination; at least there are no distinct male-specific genes in the putative sex-linked region. It is then likely that sex in catfish is determined by a very tiny difference in the genomic sequence between males and females. Such difference could be a duplicated copy of a specific gene, a single nucleotide polymorphism (SNP) locus or a combination of alleles, or a significantly differential expression pattern of a gene at the critical early embryonic stage.

Future directions

Our future study will be focus on subtle differences between female and male catfish. To identify the genes expressed exclusively in males during gonad sex differentiation period, a RNA-Seq analysis during the early embryonic developmental stage is underway. The potentially differentially expressed genes in the testis identified in the testis RNA-Seq, as well as the ones from the early embryonic RNA-Seq will eventually pinpoint the expression candidates of catfish sex determination and differentiation.

A genome-wide linkage study of sex determination in catfish using our newly developed catfish 250K SNP array would allow us to identify a set of gene loci potentially associated with sex determination. If the loci identified happen to reside with the expression candidates, further functional studies including the overexpression and knock-down of the candidate sex-determining genes need to be performed as well.

CUMULATIVE BIBLIOGRAPHY

- Affara, N.A., and Mitchell, M.J. (2000). The role of human and mouse Y chromosome genes in male infertility. *Journal of endocrinological investigation* 23, 630-645.
- Anderson, J.L., Mari, A.R., Braasch, I., Amores, A., Hohenlohe, P., Batzel, P., and Postlethwait, J.H. (2012). Multiple Sex-Associated Regions and a Putative Sex Chromosome in Zebrafish Revealed by RAD Mapping and Population Genomics. *PLoS One* 7.
- Assis, R., Zhou, Q., and Bachtrog, D. (2012). Sex-biased transcriptome evolution in *Drosophila*. *Genome Biol Evol* 4, 1189-1200.
- Bagheri-Fam, S., Sinclair, A.H., Koopman, P., and Harley, V.R. (2010). Conserved regulatory modules in the Sox9 testis-specific enhancer predict roles for SOX, TCF/LEF, Forkhead, DMRT, and GATA proteins in vertebrate sex determination. *Int J Biochem Cell Biol* 42, 472-477.
- Bak, C.W., Yoon, T.K., and Choi, Y. (2011). Functions of PIWI proteins in spermatogenesis. *Clin Exp Reprod Med* 38, 61-67.
- Bao, S., Jiang, R., Kwan, W., Wang, B., Ma, X., and Song, Y.Q. (2011). Evaluation of next-generation sequencing software in mapping and assembly. *J Hum Genet* 56, 406-414.
- Baroiller, J.F., D'Cotta, H., and Saillant, E. (2009). Environmental effects on fish sex determination and differentiation. *Sex Dev* 3, 118-135.
- Barrionuevo, F., Bagheri-Fam, S., Klattig, J., Kist, R., Taketo, M.M., Englert, C., and Scherer, G. (2006). Homozygous inactivation of Sox9 causes complete XY sex reversal in mice. *Biol Reprod* 74, 195-201.

Berbejillo, J., Martinez-Bengochea, A., Bedo, G., Brunet, F., Volff, J.N., and Vizziano-Cantonnet, D. (2012). Expression and phylogeny of candidate genes for sex differentiation in a primitive fish species, the Siberian sturgeon, *Acipenser baerii*. *Mol Reprod Dev* 79, 504-516.

Boetzer, M., Henkel, C.V., Jansen, H.J., Butler, D., and Pirovano, W. (2011). Scaffolding pre-assembled contigs using SSPACE. *Bioinformatics* 27, 578-579.

Bradley, K.M., Breyer, J.P., Melville, D.B., Broman, K.W., Knapik, E.W., and Smith, J.R. (2011). An SNP-Based Linkage Map for Zebrafish Reveals Sex Determination Loci. *G3-Genes Genom Genet* 1, 3-9.

Brunelli, J.P., Steele, C.A., and Thorgaard, G.H. (2010). Deep divergence and apparent sex-biased dispersal revealed by a Y-linked marker in rainbow trout. *Mol Phylogenet Evol* 56, 983-990.

Brunelli, J.P., Wertzler, K.J., Sundin, K., and Thorgaard, G.H. (2008). Y-specific sequences and polymorphisms in rainbow trout and Chinook salmon. *Genome* 51, 739-748.

Butler, J., MacCallum, I., Kleber, M., Shlyakhter, I.A., Belmonte, M.K., Lander, E.S., Nusbaum, C., and Jaffe, D.B. (2008). ALLPATHS: De novo assembly of whole-genome shotgun microreads. *Genome Res* 18, 810-820.

Cahais, V., Gayral, P., Tsagkogeorga, G., Melo-Ferreira, J., Ballenghien, M., Weinert, L., Chiari, Y., Belkhir, K., Ranwez, V., and Galtier, N. (2012). Reference-free transcriptome assembly in non-model animals from next-generation sequencing data. *Mol Ecol Resour* 12, 834-845.

Cao, D., Kocabas, A., Ju, Z., Karsi, A., Li, P., Patterson, A., and Liu, Z. (2001). Transcriptome of channel catfish (*Ictalurus punctatus*): initial analysis of genes and expression profiles of the head kidney. *Animal genetics* 32, 169-188.

Castano-Sanchez, C., Fuji, K., Ozaki, A., Hasegawa, O., Sakamoto, T., Morishima, K., Nakayama, I., Fujiwara, A., Masaoka, T., Okamoto, H., *et al.* (2010). A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *Bmc Genomics* 11, 554.

Castrillon, D.H., Quade, B.J., Wang, T.Y., Quigley, C., and Crum, C.P. (2000). The human VASA gene is specifically expressed in the germ cell lineage. *P Natl Acad Sci USA* 97, 9585-9590.

Chaboissier, M.C., Kobayashi, A., Vidal, V.I., Lutzkendorf, S., van de Kant, H.J., Wegner, M., de Rooij, D.G., Behringer, R.R., and Schedl, A. (2004). Functional analysis of Sox8 and Sox9 during sex determination in the mouse. *Development* 131, 1891-1901.

Chen, F., Lee, Y., Jiang, Y., Wang, S., Peatman, E., Abernathy, J., Liu, H., Liu, S., Kucuktas, H., Ke, C., and Liu, Z. (2010). Identification and characterization of full-length cDNAs in channel catfish (*Ictalurus punctatus*) and blue catfish (*Ictalurus furcatus*). *PloS one* 5, e11546.

Clemente, E.J., Furlong, R.A., Loveland, K.L., and Affara, N.A. (2006). Gene expression study in the juvenile mouse testis: identification of stage-specific molecular pathways during spermatogenesis. *Mamm Genome* 17, 956-975.

Cline, T.W., and Meyer, B.J. (1996). Vive la difference: males vs females in flies vs worms. *Annu Rev Genet* 30, 637-702.

- Cnaani, A., Lee, B.Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.F., D'Cotta, H., Penman, D.J., *et al.* (2008). Genetics of sex determination in tilapiine species. *Sex Dev* 2, 43-54.
- Conesa, A., Gotz, S., Garcia-Gomez, J.M., Terol, J., Talon, M., and Robles, M. (2005). Blast2GO: a universal tool for annotation, visualization and analysis in functional genomics research. *Bioinformatics* 21, 3674-3676.
- Cota, C.D., and Garcia-Garcia, M.J. (2012). The ENU-induced cetus mutation reveals an essential role of the DNA helicase DDX11 for mesoderm development during early mouse embryogenesis. *Dev Dynam* 241, 1249-1259.
- Davis, K.B., Simco, B.A., Goudie, C.A., Parker, N.C., Cauldwell, W., and Snellgrove, R. (1990). Hormonal Sex Manipulation and Evidence for Female Homogamety in Channel Catfish. *Gen Comp Endocr* 78, 218-223.
- de Mitcheson, Y.S., and Liu, M. (2008). Functional hermaphroditism in teleosts. *Fish Fish* 9, 1-43.
- Devlin, R.H., and Nagahama, Y. (2002). Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208, 191-364.
- Drummond, A.E. (2005). TGFbeta signalling in the development of ovarian function. *Cell Tissue Res* 322, 107-115.
- El Jamil, A., Kanhoush, R., Magre, S., Boizet-Bonhoure, B., and Penrad-Mobayed, M. (2008). Sex-Specific Expression of SOX9 During Gonadogenesis in the Amphibian *Xenopus tropicalis*. *Dev Dynam* 237, 2996-3005.

Ellegren, H., and Parsch, J. (2007). The evolution of sex-biased genes and sex-biased gene expression. *Nat Rev Genet* 8, 689-698.

Eshel, O., Shirak, A., Weller, J.I., Hulata, G., and Ron, M. (2012). Linkage and Physical Mapping of Sex Region on LG23 of Nile Tilapia (*Oreochromis niloticus*). *G3 (Bethesda)* 2, 35-42.

Fan, Y.S., Hu, Y.J., and Yang, W.X. (2012). TGF-beta superfamily: how does it regulate testis development. *Molecular Biology Reports* 39, 4727-4741.

Galindo, H.M., Loher, T., and Hauser, L. (2011). Genetic sex identification and the potential evolution of sex determination in Pacific halibut (*Hippoglossus stenolepis*). *Mar Biotechnol (NY)* 13, 1027-1037.

Gallach, M., Domingues, S., and Betran, E. (2011). Gene duplication and the genome distribution of sex-biased genes. *Int J Evol Biol* 2011, 989438.

Godbout, R., and Squire, J. (1993). Amplification of a Dead Box Protein Gene in Retinoblastoma Cell-Lines. *P Natl Acad Sci USA* 90, 7578-7582.

Graves, J.A. (2006). Sex chromosome specialization and degeneration in mammals. *Cell* 124, 901-914.

Graves, J.A.M., and Peichel, C.L. (2010). Are homologies in vertebrate sex determination due to shared ancestry or to limited options? *Genome Biol* 11.

Green, C.C., and Kelly, A.M. (2009). Effects of the estrogen mimic genistein as a dietary component on sex differentiation and ethoxyresorufin-O-deethylase (EROD) activity in channel catfish (*Ictalurus punctatus*). *Fish Physiol Biochem* 35, 377-384.

Guo, Y.Q., Cheng, H.H., Huang, X., Gao, S., Yu, H.S., and Zhou, R.J. (2005). Gene structure, multiple alternative splicing, and expression in gonads of zebrafish *Dmrt1*. *Biochem Bioph Res Co* 330, 950-957.

Hale, M.C., Xu, P., Scardina, J., Wheeler, P.A., Thorgaard, G.H., and Nichols, K.M. (2011). Differential gene expression in male and female rainbow trout embryos prior to the onset of gross morphological differentiation of the gonads. *Bmc Genomics* 12.

Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S.K., Sakamoto, T., Fernandino, J.I., Somoza, G.M., Yokota, M., and Strussmann, C.A. (2012). A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. *P Natl Acad Sci USA* 109, 2955-2959.

Herpin, A., and Scharl, M. (2011). *Dmrt1* genes at the crossroads: a widespread and central class of sexual development factors in fish. *Febs J* 278, 1010-1019.

Hosokawa, M., Shoji, M., Kitamura, K., Tanaka, T., Noce, T., Chuma, S., and Nakatsuji, N. (2007). Tudor-related proteins TDRD1/MTR-1, TDRD6 and TDRD7/TRAP: Domain composition, intracellular localization, and function in male germ cells in mice. *Dev Biol* 301, 38-52.

Howe, K., Clark, M.D., Torroja, C.F., Torrance, J., Berthelot, C., Muffato, M., Collins, J.E., Humphray, S., McLaren, K., Matthews, L., *et al.* (2013). The zebrafish reference genome sequence and its relationship to the human genome. *Nature*.

Hualde, J.P., Torres, W.D.C., Moreno, P., Ferrada, M., Demicheli, M.A., Molinari, L.J., and Luquet, C.M. (2011). Growth and feeding of Patagonian pejerrey *Odontesthes hatcheri* reared in net cages. *Aquac Res* 42, 754-763.

Huang, X., and Madan, A. (1999). CAP3: A DNA sequence assembly program. *Genome Res* 9, 868-877.

Ito, H., Fujitani, K., Usui, K., ShimizuNishikawa, K., Tanaka, S., and Yamamoto, D. (1996). Sexual orientation in *Drosophila* is altered by the satori mutation in the sex-determination gene fruitless that encodes a zinc finger protein with a BTB domain. *P Natl Acad Sci USA* 93, 9687-9692.

Ji, Y., Shi, Y.X., Ding, G.H., and Li, Y.X. (2011). A new strategy for better genome assembly from very short reads. *Bmc Bioinformatics* 12.

Jiang, T., Hou, C.C., She, Z.Y., and Yang, W.X. (2012). The SOX gene family: function and regulation in testis determination and male fertility maintenance. *Mol Biol Rep.*

Jiang, Y., Lu, J., Peatman, E., Kucuktas, H., Liu, S., Wang, S., Sun, F., and Liu, Z. (2011). A pilot study for channel catfish whole genome sequencing and de novo assembly. *Bmc Genomics* 12, 629.

Ju, Z., Karsi, A., Kocabas, A., Patterson, A., Li, P., Cao, D., Dunham, R., and Liu, Z. (2000). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): genes and expression profile from the brain. *Gene* 261, 373-382.

Kai, W., Kikuchi, K., Fujita, M., Suetake, H., Fujiwara, A., Yoshiura, Y., Ototake, M., Venkatesh, B., Miyaki, K., and Suzuki, Y. (2005). A genetic linkage map for the tiger pufferfish, *Takifugu rubripes*. *Genetics* 171, 227-238.

Kamiya, T., Kai, W., Tasumi, S., Oka, A., Matsunaga, T., Mizuno, N., Fujita, M., Suetake, H., Suzuki, S., Hosoya, S., *et al.* (2012). A Trans-Species Missense SNP in *Amhr2* Is Associated with Sex Determination in the Tiger Pufferfish, *Takifugu rubripes* (Fugu). *Plos Genet* 8.

Karsi, A., Cao, D., Li, P., Patterson, A., Kocabas, A., Feng, J., Ju, Z., Mickett, K.D., and Liu, Z. (2002). Transcriptome analysis of channel catfish (*Ictalurus punctatus*): initial analysis of gene expression and microsatellite-containing cDNAs in the skin. *Gene* 285, 157-168.

Kent, J., Wheatley, S.C., Andrews, J.E., Sinclair, A.H., and Koopman, P. (1996). A male-specific role for SOX9 in vertebrate sex determination. *Development* 122, 2813-2822.

Kikuchi, K., and Hamaguchi, S. (2013). Novel sex-determining genes in fish and sex chromosome evolution. *Dev Dyn* 242, 339-353.

Kikuchi, K., Kai, W., Hosokawa, A., Mizuno, N., Suetake, H., Asahina, K., and Suzuki, Y. (2007). The sex-determining locus in the tiger pufferfish, *Takifugu rubripes*. *Genetics* 175, 2039-2042.

Kobayashi, Y., Nagahama, Y., and Nakamura, M. (2013). Diversity and plasticity of sex determination and differentiation in fishes. *Sex Dev* 7, 115-125.

Kocabas, A.M., Li, P., Cao, D., Karsi, A., He, C., Patterson, A., Ju, Z., Dunham, R.A., and Liu, Z. (2002). Expression profile of the channel catfish spleen: analysis of genes involved in immune functions. *Mar Biotechnol (NY)* 4, 526-536.

Kojima, K., Kuramochi-Miyagawa, S., Chuma, S., Tanaka, T., Nakatsuji, N., Kimura, T., and Nakano, T. (2009). Associations between PIWI proteins and TDRD1/MTR-1 are critical for integrated subcellular localization in murine male germ cells. *Genes Cells* 14, 1155-1165.

Kondo, M., Hornung, U., Nanda, I., Imai, S., Sasaki, T., Shimizu, A., Asakawa, S., Hori, H., Schmid, M., Shimizu, N., and Scharl, M. (2006). Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. *Genome Res* 16, 815-826.

Kondo, M., Nanda, I., Hornung, U., Asakawa, S., Shimizu, N., Mitani, H., Schmid, M., Shima, A., and Scharl, M. (2003). Absence of the candidate male sex-determining gene *dmrt1b(Y)* of medaka from other fish species. *Curr Biol* 13, 416-420.

Koopman, P. (1999). *Sry* and *Sox9*: mammalian testis-determining genes. *Cellular and Molecular Life Sciences* 55, 839-856.

Koopman, P., Gubbay, J., Vivian, N., Goodfellow, P., and Lovell-Badge, R. (1991). Male development of chromosomally female mice transgenic for *Sry*. *Nature* 351, 117-121.

Kopp, A. (2012). *Dmrt* genes in the development and evolution of sexual dimorphism. *Trends Genet* 28, 175-184.

Kucuktas, H., Wang, S., Li, P., He, C., Xu, P., Sha, Z., Liu, H., Jiang, Y., Baoprasertkul, P., Somridhivej, B., *et al.* (2009). Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. *Genetics* 181, 1649-1660.

Leder, E.H., Cano, J.M., Leinonen, T., O'Hara, R.B., Nikinmaa, M., Primmer, C.R., and Merila, J. (2010). Female-Biased Expression on the X Chromosome as a Key Step in Sex Chromosome Evolution in Threespine Sticklebacks. *Mol Biol Evol* 27, 1495-1503.

Lee, B.Y., Coutanceau, J.P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J.F., and Kocher, T.D. (2011). Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol (NY)* 13, 557-562.

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

Li, P., Peatman, E., Wang, S., Feng, J., He, C., Baoprasertkul, P., Xu, P., Kucuktas, H., Nandi, S., Somridhivej, B., *et al.* (2007). Towards the ictalurid catfish transcriptome: generation and analysis of 31,215 catfish ESTs. *BMC genomics* 8, 177.

Li, R., Zhu, H., Ruan, J., Qian, W., Fang, X., Shi, Z., Li, Y., Li, S., Shan, G., Kristiansen, K., *et al.* (2010). De novo assembly of human genomes with massively parallel short read sequencing. *Genome Res* 20, 265-272.

Liew, W.C., Bartfai, R., Lim, Z., Sreenivasan, R., Siegfried, K.R., and Orban, L. (2012). Polygenic sex determination system in zebrafish. *PLoS One* 7, e34397.

Liu, L., Li, Y., Li, S., Hu, N., He, Y., Pong, R., Lin, D., Lu, L., and Law, M. (2012a). Comparison of next-generation sequencing systems. *J Biomed Biotechnol* 2012, 251364.

Liu, S., Zhang, Y., Zhou, Z., Waldbieser, G., Sun, F., Lu, J., Zhang, J., Jiang, Y., Zhang, H., Wang, X., *et al.* (2012b). Efficient assembly and annotation of the transcriptome of catfish by RNA-Seq analysis of a doubled haploid homozygote. *Bmc Genomics* 13, 595.

Liu, S., Zhou, Z., Lu, J., Sun, F., Wang, S., Liu, H., Jiang, Y., Kucuktas, H., Kaltenboeck, L., Peatman, E., and Liu, Z. (2011). Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC genomics* 12, 53.

- Liu, Z.H., Zhang, Y.G., and Wang, D.S. (2010). Studies on feminization, sex determination, and differentiation of the Southern catfish, *Silurus meridionalis*-a review. *Fish Physiol Biochem* 36, 223-235.
- Liu, Z.J. (2003). A review of catfish genomics: progress and perspectives. *Comp Funct Genom* 4, 259-265.
- MacCallum, I., Przybylski, D., Gnerre, S., Burton, J., Shlyakhter, I., Gnirke, A., Malek, J., McKernan, K., Ranade, S., Shea, T.P., *et al.* (2009). ALLPATHS 2: small genomes assembled accurately and with high continuity from short paired reads. *Genome Biol* 10.
- Marchand, O., Govoroun, M., D'Cotta, H., McMeel, O., Lareyre, J.J., Bernot, A., Laudet, V., and Guiguen, Y. (2000). DMRT1 expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. *Bba-Gene Struct Expr* 1493, 180-187.
- Mardis, E.R. (2008). Next-generation DNA sequencing methods. *Annu Rev Genomics Hum Genet* 9, 387-402.
- Marguerat, S., and Bahler, J. (2010). RNA-seq: from technology to biology. *Cell Mol Life Sci* 67, 569-579.
- Margulies, M., Egholm, M., Altman, W.E., Attiya, S., Bader, J.S., Bemben, L.A., Berka, J., Braverman, M.S., Chen, Y.J., Chen, Z., *et al.* (2005). Genome sequencing in microfabricated high-density picolitre reactors. *Nature* 437, 376-380.
- Martinez, P., Bouza, C., Hermida, M., Fernandez, J., Toro, M.A., Vera, M., Pardo, B., Millan, A., Fernandez, C., Vilas, R., *et al.* (2009). Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). *Genetics* 183, 1443-1452.

Matsuda, M., Nagahama, Y., Shinomiya, A., Sato, T., Matsuda, C., Kobayashi, T., Morrey, C.E., Shibata, N., Asakawa, S., Shimizu, N., *et al.* (2002). DMY is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* *417*, 559-563.

Matsuda, M., Shinomiya, A., Kinoshita, M., Suzuki, A., Kobayashi, T., Paul-Prasanth, B., Lau, E.L., Hamaguchi, S., Sakaizumi, M., and Nagahama, Y. (2007). DMY gene induces male development in genetically female (XX) medaka fish. *Proc Natl Acad Sci U S A* *104*, 3865-3870.

Miller, J.R., Koren, S., and Sutton, G. (2010). Assembly algorithms for next-generation sequencing data. *Genomics* *95*, 315-327.

Mishina, Y., Rey, R., Finegold, M.J., Matzuk, M.M., Josso, N., Cate, R.L., and Behringer, R.R. (1996). Genetic analysis of the Mullerian-inhibiting substance signal transduction pathway in mammalian sexual differentiation. *Genes Dev* *10*, 2577-2587.

Morozova, O., and Marra, M.A. (2008). Applications of next-generation sequencing technologies in functional genomics. *Genomics* *92*, 255-264.

Myosho, T., Otake, H., Masuyama, H., Matsuda, M., Kuroki, Y., Fujiyama, A., Naruse, K., Hamaguchi, S., and Sakaizumi, M. (2012). Tracing the Emergence of a Novel Sex-Determining Gene in Medaka, *Oryzias luzonensis*. *Genetics* *191*, 163-+.

Nakamoto, M., Suzuki, A., Matsuda, M., Nagahama, Y., and Shibata, N. (2005). Testicular type Sox9 is not involved in sex determination but might be in the development of testicular structures in the medaka, *Oryzias latipes*. *Biochem Biophys Res Commun* *333*, 729-736.

Nakamura, M., Kobayashi, T., Chang, X.T., and Nagahama, Y. (1998). Gonadal sex differentiation in teleost fish. *J Exp Zool* *281*, 362-372.

Nanda, I., Kondo, M., Hornung, U., Asakawa, S., Winkler, C., Shimizu, A., Shan, Z., Haaf, T., Shimizu, N., Shima, A., *et al.* (2002). A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. *Proc Natl Acad Sci U S A* *99*, 11778-11783.

Ninwichian, P., Peatman, E., Liu, H., Kucuktas, H., Somridhivej, B., Liu, S., Li, P., Jiang, Y., Sha, Z., Kaltenboeck, L., *et al.* (2012a). Second-generation genetic linkage map of catfish and its integration with the BAC-based physical map. *G3 (Bethesda)* *2*, 1233-1241.

Ninwichian, P., Peatman, E., Perera, D., Liu, S., Kucuktas, H., Dunham, R., and Liu, Z. (2012b). Identification of a sex-linked marker for channel catfish. *Anim Genet* *43*, 476-477.

Parisi, M., Nuttall, R., Naiman, D., Bouffard, G., Malley, J., Andrews, J., Eastman, S., and Oliver, B. (2003). Paucity of genes on the *Drosophila* X chromosome showing male-biased expression. *Science* *299*, 697-700.

Paszkievicz, K., and Studholme, D.J. (2010). De novo assembly of short sequence reads. *Briefings in Bioinformatics* *11*, 457-472.

Patino, R., Davis, K.B., Schoore, J.E., Uguz, C., Strussmann, C.A., Parker, N.C., Simco, B.A., and Goudie, C.A. (1996). Sex differentiation of channel catfish gonads: Normal development and effects of temperature. *J Exp Zool* *276*, 209-218.

Peichel, C.L., Ross, J.A., Matson, C.K., Dickson, M., Grimwood, J., Schmutz, J., Myers, R.M., Mori, S., Schluter, D., and Kingsley, D.M. (2004). The master sex-determination locus in threespine sticklebacks is on a nascent Y chromosome. *Curr Biol* *14*, 1416-1424.

- Piferrer, F., and Guiguen, Y. (2008). Fish Gonadogenesis. Part II: Molecular Biology and Genomics of Sex Differentiation. *Rev Fish Sci* 16, 35-55.
- Pop, M. (2009). Genome assembly reborn: recent computational challenges. *Brief Bioinform* 10, 354-366.
- Pop, M., Phillippy, A., Delcher, A.L., and Salzberg, S.L. (2004). Comparative genome assembly. *Brief Bioinform* 5, 237-248.
- Puri, P., Snow, A.J., Kline, D.W., and Vijayaraghavan, S. (2008). Identification of 14-3-3 Binding Proteins in Mouse Testis by Tandem Affinity Tag Purification. *Faseb J* 22.
- Quiniou, S.M., Waldbieser, G.C., and Duke, M.V. (2007). A first generation BAC-based physical map of the channel catfish genome. *Bmc Genomics* 8, 40.
- Raghuveer, K., and Senthilkumaran, B. (2009). Identification of multiple dmrt1s in catfish: localization, dimorphic expression pattern, changes during testicular cycle and after methyltestosterone treatment. *J Mol Endocrinol* 42, 437-448.
- Raghuveer, K., Senthilkumaran, B., Sudhakumari, C.C., Sridevi, P., Rajakumar, A., Singh, R., Muruganankumar, R., and Majumdar, K.C. (2011). Dimorphic Expression of Various Transcription Factor and Steroidogenic Enzyme Genes during Gonadal Ontogeny in the Air-Breathing Catfish, *Clarias gariepinus*. *Sexual Development* 5, 213-223.
- Robertson, G., Schein, J., Chiu, R., Corbett, R., Field, M., Jackman, S.D., Mungall, K., Lee, S., Okada, H.M., Qian, J.Q., *et al.* (2010). De novo assembly and analysis of RNA-seq data. *Nat Methods* 7, 909-U962.

Rolland, A.D., Lareyre, J.J., Goupil, A.S., Montfort, J., Ricordel, M.J., Esquerre, D., Hugot, K., Houlgatte, R., Chalmel, F., and Le Gac, F. (2009). Expression profiling of rainbow trout testis development identifies evolutionary conserved genes involved in spermatogenesis. *Bmc Genomics* 10.

Rothberg, J.M., and Leamon, J.H. (2008). The development and impact of 454 sequencing. *Nat Biotechnol* 26, 1117-1124.

Saccone, G., Pane, A., and Polito, L.C. (2002). Sex determination in flies, fruitflies and butterflies. *Genetica* 116, 15-23.

Saffman, E.E., and Lasko, P. (1999). Germline development in vertebrates and invertebrates. *Cellular and Molecular Life Sciences* 55, 1141-1163.

Sarre, S.D., Ezaz, T., and Georges, A. (2011). Transitions Between Sex-Determining Systems in Reptiles and Amphibians. *Annu Rev Genom Hum G* 12, 391-406.

Shao, C.W., Chen, S.L., Scheuring, C.F., Xu, J.Y., Sha, Z.X., Dong, X.L., and Zhang, H.B. (2010). Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. *Mar Biotechnol (NY)* 12, 558-568.

Shapiro, M.D., Summers, B.R., Balabhadra, S., Aldenhoven, J.T., Miller, A.L., Cunningham, C.B., Bell, M.A., and Kingsley, D.M. (2009). The genetic architecture of skeletal convergence and sex determination in ninespine sticklebacks. *Curr Biol* 19, 1140-1145.

Shendure, J., and Ji, H. (2008). Next-generation DNA sequencing. *Nat Biotechnol* 26, 1135-1145.

Shirak, A., Seroussi, E., Cnaani, A., Howe, A.E., Domokhovskiy, R., Zilberman, N., Kocher, T.D., Hulata, G., and 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.

Simpson, J.T., Wong, K., Jackman, S.D., Schein, J.E., Jones, S.J.M., and Birol, I. (2009). ABySS: A parallel assembler for short read sequence data. *Genome Res* *19*, 1117-1123.

Sinclair, A.H., Berta, P., Palmer, M.S., Hawkins, J.R., Griffiths, B.L., Smith, M.J., Foster, J.W., Frischauf, A.M., Lovell-Badge, R., and Goodfellow, P.N. (1990). A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. *Nature* *346*, 240-244.

Small, C.M., Carney, G.E., Mo, Q.X., Vannucci, M., and Jones, A.G. (2009). A microarray analysis of sex- and gonad-biased gene expression in the zebrafish: Evidence for masculinization of the transcriptome. *Bmc Genomics* *10*.

Smith, C.A., Roeszler, K.N., Ohnesorg, T., Cummins, D.M., Farlie, P.G., Doran, T.J., and Sinclair, A.H. (2009). The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. *Nature* *461*, 267-271.

Sun, F., Peatman, E., Li, C., Liu, S., Jiang, Y., Zhou, Z., and Liu, Z. (2012). Transcriptomic signatures of attachment, NF-kappaB suppression and IFN stimulation in the catfish gill following columnaris bacterial infection. *Dev Comp Immunol* *38*, 169-180.

Takase, M., Noguchi, S., and Nakamura, M. (2000). Two Sox9 messenger RNA isoforms: isolation of cDNAs and their expression during gonadal development in the frog *Rana rugosa*. *Febs Lett* 466, 249-254.

Tanaka, T., Hosokawa, M., Vagin, V.V., Reuter, M., Hayashi, E., Mochizuki, A.L., Kitamura, K., Yamanaka, H., Kondoh, G., Okawa, K., *et al.* (2011). Tudor domain containing 7 (Tdrd7) is essential for dynamic ribonucleoprotein (RNP) remodeling of chromatoid bodies during spermatogenesis. *P Natl Acad Sci USA* 108, 10579-10584.

Tiersch, T.R., Simco, B.A., Davis, K.B., Chandler, R.W., Wachtel, S.S., and Carmichael, G.J. (1990). Stability of Genome Size among Stocks of the Channel Catfish. *Aquaculture* 87, 15-22.

Tiersch, T.R., Simco, B.A., Davis, K.B., and Wachtel, S.S. (1992). Molecular genetics of sex determination in channel catfish: studies on SRY, ZFY, Bkm, and human telomeric repeats. *Biol Reprod* 47, 185-192.

Tripathi, N., Hoffmann, M., Weigel, D., and Dreyer, C. (2009). Linkage Analysis Reveals the Independent Origin of Poeciliid Sex Chromosomes and a Case of Atypical Sex Inheritance in the Guppy (*Poecilia reticulata*). *Genetics* 182, 365-374.

Vidal, V.P., Chaboissier, M.C., de Rooij, D.G., and Schedl, A. (2001). Sox9 induces testis development in XX transgenic mice. *Nat Genet* 28, 216-217.

Vinas, A., Taboada, X., Vale, L., Robledo, D., Hermida, M., Vera, M., and Martinez, P. (2012). Mapping of DNA sex-specific markers and genes related to sex differentiation in turbot (*Scophthalmus maximus*). *Mar Biotechnol (NY)* 14, 655-663.

Waldbieser, G.C., Bosworth, B.G., Nonneman, D.J., and Wolters, W.R. (2001). A microsatellite-based genetic linkage map for channel catfish, *Ictalurus punctatus*. *Genetics* 158, 727-734.

Waldbieser, G.C., Bosworth, B.G., and Quiniou, S.M. (2010). Production of viable homozygous, doubled haploid channel catfish (*Ictalurus punctatus*). *Mar Biotechnol (NY)* 12, 380-385.

Wang, F., Yu, Y., Ji, D.R., and Li, H.Y. (2012). The DMRT gene family in amphioxus. *J Biomol Struct Dyn* 30, 191-200.

Wang, S., Peatman, E., Abernathy, J., Waldbieser, G., Lindquist, E., Richardson, P., Lucas, S., Wang, M., Li, P., Thimmapuram, J., *et al.* (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, R8.

Wilhelm, D., Palmer, S., and Koopman, P. (2007). Sex determination and gonadal development in mammals. *Physiol Rev* 87, 1-28.

Xu, P., Wang, S., Liu, L., Thorsen, J., Kucuktas, H., and Liu, Z. (2007). A BAC-based physical map of the channel catfish genome. *Genomics* 90, 380-388.

Yano, A., Guyomard, R., Nicol, B., Jouanno, E., Quillet, E., Klopp, C., Cabau, C., Bouchez, O., Fostier, A., and Guiguen, Y. (2012). An Immune-Related Gene Evolved into the Master Sex-Determining Gene in Rainbow Trout, *Oncorhynchus mykiss*. *Curr Biol* 22, 1423-1428.

Yoshimoto, S., and Ito, M. (2011). A ZZ/ZW-type sex determination in *Xenopus laevis*. *Febs J* 278, 1020-1026.

Yoshimoto, S., Okada, E., Umemoto, H., Tamura, K., Uno, Y., Nishida-Umehara, C., Matsuda, Y., Takamatsu, N., Shiba, T., and Ito, M. (2008). A W-linked DM-domain gene, DM-W,

participates in primary ovary development in *Xenopus laevis*. *Proc Natl Acad Sci U S A* *105*, 2469-2474.

Zerbino, D.R., and Birney, E. (2008). Velvet: Algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res* *18*, 821-829.

Zerbino, D.R., McEwen, G.K., Margulies, E.H., and Birney, E. (2009). Pebble and Rock Band: Heuristic Resolution of Repeats and Scaffolding in the Velvet Short-Read de Novo Assembler. *PLoS One* *4*.