TUMSAT-OACIS Repository - Tokyo

University of Marine Science and Technology

(東京海洋大学)

Genomics and genetics of flatfishes provide insights into the mechanisms of sex determination and disease resistance

メタデータ	言語: eng 出版者: 公開日: 2015-06-24 キーワード (Ja): キーワード (En): 作成者: Shao, Changwei メールアドレス: 所属:
URL	https://oacis.repo.nii.ac.jp/records/1085

Doctoral Dissertation

GENOMICS AND GENETICS OF FLATFISHES PROVIDE INSIGHTS INTO THE MECHANISMS OF SEX DETERMINATION AND DISEASE RESISTANCE

異体類のゲノム科学と遺伝学は性決定と耐病性のメカニズムの解明に手掛かりをもたらす

March 2015

Changwei Shao

Sometimes I feel weighed down as a researcher. Sometimes I feel overwhelmed by grant deadlines, unwritten papers, finicky experiments, endless forms, and so forth. Why am I doing this? Isn't there an easier way to earn a living? During times like this, I'll sometimes glance over to the corner of my desk. There I display pictures of special people and things — my mother, my wife, my little daughter … and a DNA sequence. Now, pictures of family are understandable, but why a DNA sequence?

Introduced from Harvey J. Armbrecht

Doctoral Dissertation

GENOMICS AND GENETICS OF FLATFISHES PROVIDE INSIGHTS INTO THE MECHANISMS OF SEX DETERMINATION AND DISEASE RESISTANCE

異体類のゲノム科学と遺伝学は性決定と耐病性のメカニズムの解明に手掛かりをもたらす

March 2015

Changwei Shao

［課程博士・論文博士共通］

博士学位論文内容要旨
Abstract

専　攻 Major		氏　名 Name	Shao Changwei
論文題目 Title	Genomics and genetics of flatfishes provide insights into the mechanisms of sex determination and disease resistance 異体類のゲノム科学と遺伝学は性決定と耐病性のメカニズムの解明に手掛かりをもたらす		

Flatfishes, members of the order Pleuronectiformes, comprise a biologically interesting and commercially important species, with a unique asymmetric body shape developed as an adaptation to a lifestyle living on the sea bottom. As a commercially important cultured flatfish, half-smooth tongue sole (*Cynoglossus semilaevis*) have a great growth dimorphosism that females grow faster and attain larger size than males. Besides the tongue sole possessed a ZZ/ZW genetic sex determination system and the functional sex reversal can be induced by environmental factors. Therefore, the coexistence of GSD and ESD system within the tongue sole, coupled with the availability of its genomic and genetic information, provides an opportunity to comprehensively investigate the mechanism of sex determination and differentiation of this species. Besides, another target flatfish species, Japanese flounder (*Paralichthys olivaceus*) is also one of the most desirable and highly priced marine fish species. With extensive cultivation, however, farming of the Japanese flounder has been confronted with certain problems, including a high mortality rate as well as a decline in growth. To further increase profitability and sustainability while maintaining genetic variability in the cultured stock, it is urgent to carry out both classic selective breeding and molecular marker assisted selection. In this thesis, a multi-level of genomic and genetic analysis focuses on the sex determination and differentiation as well as the disease resistance of flatfishes were carried out. The main results are as follows:

Two bacterial artificial chromosome (BAC) libraries for tongue sole were constructed in the *Bam*HI and *Hind*III sites of the vector pECBAC1. The two libraries contain a total of 55,296 BAC clones with average insert size of approximately 156.4 Kb and correspond to 13.36 haploid genome equivalents. The combined libraries have a greater than 99% probability of containing any single-copy sequence. Screening high-density arrays of the libraries generated 76 and 152 positive clones with average of 15.2 and 30 per probe for sex-related genes and female-specific markers. Five positive BAC clones containing sex-related genes and female-specific markers were sequenced by ABI 3730 DNA analyzer. A full-length BAC sequence with nine genes including *cyp19a1a* gene was further analyzed on the structural characteristics such as repeats, GC content, transposable elements, etc. A comparative analysis on the *cyp19a1a* gene was carried out among medaka, zebrafish, three-spine stickleback, *Fugu rubripes* and *Tetraodon*, suggesting that the conservation of *cyp19a1a* gene structure and function, mainly including the same pattern of exons and introns, five conserved functional domains, the common upstream regulatory elements and the high expression in ovarian specificity for all the teleost *cyp19a1a* gene.

The single-base-resolution methylomes for tongue sole were depicted using the whole genome bisulfite

sequencing (WGBS). In total, 171 Gb methylome data were produced, which yielded an average depth of 22× per strand for adult gonads of parental females (P-ZWf), parental pseudomales (P-ZWm), F1 pseudomales (F1-ZWm), and females (F1-ZWf) as well as the normal male (ZZm) as a control. The results show that the methylation patterns of pseudomale testes were similar to those of normal male testes, and all three-testis samples were clearly distinguished from the ovary samples by hierarchical clustering analysis. Only 60 Kb of differentially methylated regions (DMRs) were identified between P-ZWf and F1-ZWf. In contrast, an average of 15 Mb of DMRs was identified between testes and ovaries. Interestingly, up to 86% of the DMRs between P-ZWm and ovaries were maintained in F1-ZWm, indicating that the overall change in methylation status of the genome after sex reversal had been inherited. Besides, 2,986 differentially methylated genes (DMGs) that harbored testis/ovary DMRs on their promoter regions were identified. In total, 16 of 58 (28%) sex-determination-related genes in tongue sole displayed significantly differential methylation patterns between testes and ovaries, which contrasts to 14% (2,986/21,516) over the whole genome, suggest that the sex determination pathway is the target of DNA methylation regulation. Transcriptome analysis revealed that the dosage compensation occurs in a restricted, methylated cytosine enriched Z chromosomal region in pseudomale testes, achieving equal expression level in normal male testes. In contrast, female-specific W chromosomal genes are suppressed in pseudomales by methylation regulation.

A total of 12,712 high-confidence SNPs were successfully high-throughput genotyped and then were assigned to the 24 consensus linkage groups (LGs) based on the restriction-site associated DNA (RAD) sequencing in Japanese flounder. The total length of the genetic linkage map was 3497.29 cM with an average distance of 0.48 cM between loci, thereby representing the dentist genetic map in Japanese flounder. Ten quantitative trait loci (QTLs) forming two main clusters for *vibrio anguillarum* disease were detected. All QTLs with logarithm of odds (LOD) value ranged from 3 to 15.8 can explain the total of phenotypic variance from 1.19% to 8.38%. By synteny analysis of the QTLs regions with genome assembly, twelve immune related genes were identified and among them, four genes involved in the function of MHC were strong associated with *V. anguillarum* disease. Furthermore, 246 genome assembly scaffolds with an average size of 21.79 Mb, accounting for 522.995 Mb totally, were anchored onto the LGs, representing 95.8% of the assembled genomic sequences of the Japanese flounder. The mapped assembly scaffolds in Japanese flounder were then used for genome synteny analysis against the zebrafish and medaka, respectively. The flounder and medaka have almost one-to-one synteny, while the flounder and zebrafish have multi-synteny correspondence, indicated that the medaka has a closer relationship with flounder.

CONTENT

CHAPTER Ⅰ INTRODUCTION 1

1. SEX DETERMINATION AND SEX CHROMOSOME EVOLUTION 1

1.1 SEX DETERMINATION SYSTEMS 1

1.2 SEX DETERMINATION GENES IN VERTEBRATES 1

1.3 CO-EVOLUTION OF SEX DETERMINATION GENE AND SEX CHROMOSOME EVOLUTION 2

1.4 MOLECULAR MECHANISM OF SEX REVERSAL 3

2. DISEASE RESISTANCE MECHANISM AND MOLECULAR ASSISTANT BREEDING 4

2.1 EVOLUTIONARY PERSPECTIVE ON MOLECULAR MECHANISM OF DISEASE RESISTANCE 4

2.2 MOLECULAR-ASSISTED BREEDING TECHNIQUES 4

2.3 ADVANCE OF FISH DISEASE RESISTANCE 5

3. GENOMIC AND GENETIC TECHNIQUES 6

3.1 NEXT GENERATION SEQUENCING TECHNIQUES 6

3.2 BAC LIBRARIES CONSTRUCTION 7

3.3 RESTRICTION-SITE ASSOCIATED DNA SEQUENCING 7

3.4 WHOLE GENOME BISULFITE SEQUENCING 8

3.5 WHOLE GENOME SEQUENCE ON TELEOST SPECIES 9

4. FLATFISH SPECIES AND OUT TARGETS 11

REFERENCES: 13

CHAPTER Ⅱ BAC LIBRARY CONSTRUCTION AND APPLICATION IN HALF-SMOOTH TONGUE SOLE 20

CHAPTER Ⅲ EPIGENETIC INHERITANCE OF SEX REVERSAL IN HALF-SMOOTH TONGUE SOLE 64

CHAPTER Ⅳ HIGH-RESOLUTION GENETIC MAP CONSTRUCTION AND ITS APPLICATION IN JAPANESE FLOUNDER 77

CONCLUSION 88

ACKNOWLEDGEMENTS 89

Chapter Ⅰ Introduction

1. Sex determination and sex chromosome evolution

1.1 Sex determination systems

Sex determination is a plastic biological developmental process, which has always the intriguing aspect in evolutionary biology and developmental biology. In vertebrates, the sex determination including genetic sex determination (GSD) and environmental sex determination (ESD) acts to differentiate an initially bipotential gonad primordium into either testes or ovaries (Marshall Graves, 2008). GSD involves the inheritance of specific elements at the genotypic level, which cause a zygote to develop into a particular sexual fate. The mechanism of GSD was divided into two types. One is the male heterogamety that all males are XY and females are XX. Another is the reverse of male heterogamety where females are ZW and males are ZZ (Namekawa and Lee, 2009). The mechanism of ESD was determined by environment factors such as temperature, social environment and hormones after fertilization. This mechanism is known as temperature-dependant sex determination (TSD). Controversy has surrounded the issue of the abundance of TSD species, and in fact species that exhibit TSD under natural conditions of temperature are less common than initially thought (Ospina-Alvarez and Piferrer, 2008). Among vertebrates, teleost species have the highest variability of sex determining mechanisms ranging from GSD system to ESD system and GSD+ESD controlled mechanisms (Stelkens and Wedekind, 2010). Therefore the teleost with such polymorphism of sex determination as a good research model to study on sex determination and differentiation as well as sex chromosome evolution.

1.2 Sex determination genes in vertebrates

GSD is governed by a series of sex-related genes involving in genetic pathway that initiate by sex-determining gene during critical periods of gonadal development. Though many downstream genes in sex determination pathways are conserved, even among vertebrates and invertebrates, the upstream sex determination gene can vary even between closely related species (Shao and Chen, 2012). Up to now, eight sex-determining genes (*sry*, *dmrt1*, *dmy*, *dmw*, *amhy*, *gsdf*, *amhr2* and *sdY*) had been identified in vertebrates. Humans and most other mammals have a strictly male heterogamety with XY for male and XX for female. The smaller Y chromosome bears the master sex determination gene, *sry* (Sinclair et al., 1990). It acts as a transcriptional factor to activate the downstream gene in the gene pathway, which is the most clear sex determination pathway to date (Waters et al., 2007). In the chicken and all other birds, the homogametic sex is male (ZZ) and the heterogametic sex is female (ZW). The candidate sex determination gene for avain is *dmrt1*, which is a conserved Z-linked gene (Smith et al., 2009). The amphibians have two systems of sex chromosomes: one with heterogametic male (XX/XY) and another with the female (ZZ/ZW). A W-linked sex determination gene, *dmw* was isolated from *Xenopus laevis*, which employed a sex determination system of ZZ/ZW (Yoshimoto et al., 2008). Teleost

fishes (over 30,000 species) are the largest group of vertebrates, which exhibit kinds of sex determination mechanism. The sex of medaka is genetically determined by a XX-XY male heterogametic system and crossing-over between the sex chromosomes is possible. Classical genetic approaches and candidate gene analyses have led to the identification of *dmy* in a short region on the Y chromosome as a male sex determining gene, which is the first reported sex determination gene in teleost (Matsuda et al., 2002; Nanda et al., 2002). Recently four novel candidates for vertebrate SD genes including the *amhy*, *gsdf*, *amhr2* and *sdY* were reported, all of them are in fishes. *Gsdf* is a strong candidate for the master SD gene in *Oryzias luzonensis*, which diverged from *O. curvinotus* with a sex determination gene of *dmy* approximately 10 Mya (Myosho et al., 2012). As a male-specific, duplicated copy of the anti-Mullerian hormone gene (*amh*), *amhy* is a candidate for the master SD gene in the *Patagonian pejerrey*, a teleost species, which possesses an XX/XY sex determination system (Hattori et al., 2012). *Amhr2* coding for a receptor of *amh* was perfectly associated with the phenotypic sex in fugu, suggesting that it is responsible for sex determination in fugu (Kamiya et al., 2012). The *sdY* gene was shown to be tightly linked to the SD locus in rainbow trout with an XX/XY system (Yano et al., 2012). It was considered as a first sex determination gene that was not followed the expected evolutionary plasticity in the mechanisms underlying vertebrate sex determination.

1.3 Co-evolution of sex determination gene and sex chromosome evolution

Generally, the sex chromosomes are doomed to be degenerate to some extent driving by the evolutionary force of suppressed recombination, which was firstly appeared around sex determination gene (Marshall Graves, 2008). Given all the sex determination genes from the fish species mentioned above, together with the sex determination genes in mammalian, bird and amphibians species, we can conclude that the sex-determining genes (*sry* and *dmrt1*) are relatively conserved in higher vertebrate, therian and aves, respectively, comparing to the sex-determining genes *dmy*, *dmw*, *amhy*, *gsdf*, *amhr2* and *sdY* which are unstable in their respective taxonomic systems. Correspondingly, the sex chromosomes in higher vertebrate attained high differentiation during their evolution, whereas no obvious (or even no) differentiation was observed in the sex chromosome of most extant lower vertebrates. The appearance of sex-determining gene in certain organism was always accompanied with the evolution of their sex chromosomes, which was defined by coevolution. In consequently, the difference of conservation on sex-determining genes between higher vertebrates and lower vertebrates may be caused by the sex chromosome differentiation or not. Following these conclusions, two models on the relationship between sex-determining gene and sex chromosome evolution were established. One is the model of differentiation on sex chromosome that developed through diversification of one region of the progenitor chromosomes. In this model, transposons and repetitive elements were accumulated around a sex-determining gene and its homolog evolved into a pseudo-gene because the recombination was ceased in the sex-determining region. In this case, a dosage dependent sex-determining gene in homogametic sex or a heteromorphic chromosome-linked (usually Y and W) determining gene in heterogametic sex such as

dmrt1 in chicken and *sry* in human were expected, respectively. The other model is undifferentiation of sex chromosome where sex-determining gene derived from a duplicated region from elsewhere in the genome that was inserted on it. The large region of sex chromosomes in such case can still recombine, though the duplicated region could accumulate few repetitive elements. In this case, the sex-determining genes would reside on the heterogametic chromosome (usually Y and W) such as *dmy*, *gsdf*, *dmw* and *amhy* (Shao and Chen, 2012). In all, sex determination is a complex biological process; these two models cannot represent the types of sex determination in all species, but above models proposed would make for the isolation of new sex-determining gene.

1.4 Molecular mechanism of sex reversal

In general, the GSD species in which sex ratios are established by a combination of genetic and environmental influences are referred to as GSD + ESD species, which means the primary sex can be altered during development under environmental factors. This great plasticity of the sex determining mechanisms has been linked to the great diversity of species in fish since plasticity in sex determination favors speciation (Qvarnström and Bailey, 2009). This phenomenon is known as environmental sex reversal (ESR) (Stelkens and Wedekind, 2010), which can lead to widely varying and unstable population sex ratios. This indicates that environmental factors can directly or indirectly override sex chromosomes or sex genes in ESR species. Those environmental factors include the abiotic (e.g. temperature, pH, endocrine-disrupting hormones, photoperiod, hypoxia) and biotic factors (crowding, pathogens like Wolbachia, population size) (Stelkens and Wedekind, 2010). Actually, in order to obtain the preferred sex in aquaculture, the exogenous steroids or high temperature was used to artificially induced sex reversal in fish. To date, more than 50 fish species have been studied in such steroid- or temperature- mediated sex reversal trials (Devlin and Nagahama, 2002; Cnaani and Levavi-Sivan, 2009). Although thermal or chemical sensitivity of gonadal sex differentiation has been documented on different fish species, the molecular mechanism and consequences of ESR are not yet well understood (Devlin and Nagahama, 2002; Baroiller et al., 2009). It should be noted here that even in species with ESR, selection is ultimately acting on genes that create a link between offspring sex and a certain environmental factor. In consequence, the analysis of epigenetic marker that significantly contributes to changes in gene expression in response to varying environments was introduced to research on sex differentiation. One important insight from recent work on the relationship between *cyp19a1a* gene expression and methylation level involving in the gonad differentiation seems to be a well example in European sea bass. The results show that the promoter of female benefit gene, *cyp19a1a*, was hypermethylated in males when compared to females, and the high temperature increased the methylation levels of *cyp19a1a* in both sexes, suggesting that an epigenetic mechanism may be involved in environmental temperature induced change on gene expression (Navarro-Martin et al., 2011).

2. Disease resistance mechanism and molecular assistant breeding

2.1 Evolutionary perspective on molecular mechanism of disease resistance

Traditional selection based on the phenotype has created a wide diversity of strains or breeds of domestic animal that are adapted to different environments and are required to feel comfortable for breeding purposes (Jovanović et al., 2009). Such phenotype selection was in nature identification of the breeds with beneficial genetic mutation that occurs because of nature selection over many generations and in large populations. Genetic resistance to domestic animal disease has been recognised as an important kind of breeding purpose. Such beneficial genetic mutation for disease resistance is difficult to select by traditional breeding programmes (Pieter and Stephen, 2000). As a result, there is a considerable interest in finding genetic markers that can be applied in breeding programmes. However, it has only been fitfully exploited because of the limitation of genetic resource as well as, in particularly, the trait of multifactorial genetics of disease that controlled by an unknown number of quantitative trait loci (QTLs). From a personal perspective, the collection of more accurate data on disease resistance and the development of genomic approaches are particularly useful for improvement of disease resistance in farm animals. Special attention has been focused on whole genome sequencing and re-sequencing for detection of genetic variation in terms of the complex responses to disease and in consequently the individuals with the beneficial genetic mutation for disease resistance would be selected. Due to their genetic advantages when in responding to disease, the maintenance of favorable genetic mutations in domestic animal populations is crucial to maintaining breeding practices and the preservation of biodiversity for domestic animals (Medugorac et al., 2011).

2.2 Molecular-assisted breeding techniques

Molecular-assisted breeding was defined by using of genetic manipulation performed at DNA molecular levels to improve characters of interest in plants and animals, including genetic engineering or gene manipulation, molecular marker-assisted selection, genome-wide association study, genomic selection, etc (Collard and Mackill, 2008). The discovery of the structure of DNA has enhanced traditional breeding techniques by allowing breeders to pinpoint the particular gene responsible for a particular trait (Xu, 2009). And thus the technique for isolating one or more specific genes and introducing them into a species was defined as gene manipulation. However, it is difficult to extend to most of economical species especially the marine species, although improvement of some species can be achieved by this technique. More often now, molecular-assisted breeding was defined to improve plant or animal traits on the basis of genotypic assays by molecular markers such as restriction fragment length polymorphism (RFLP), randomly amplified polymorphic DNA (RAPD), simple sequence repeats (SSRs) and single nucleotide polymorphism (SNP). Because of their abundance and importance in the genome, molecular markers have been widely used in the fields of genetic mapping, marker-assisted breeding, genome-wide association study and genomic selection. Molecular marker

technology has become a powerful tool in the genetic breeding. In particular, as a marker-assisted breeding technique, the QTL mapping was extensively used for various economical species and the associated QTLs for kinds of traits were identified (Wurschum, 2012). However, the application of QTLs to the breeding is difficult because it cannot identify all trait-associated genes and transcript variants. Recently, the genome-wide association studies have successfully identified numerous loci at which common variants influence human disease risk (Lettre and Rioux, 2008). But it is still difficult for quantitative traits of plant and animals. The genomic selection, of particularly, a new approach for improving quantitative traits, were attained widely application on a population that is different from the reference population in which the marker effects were estimated. Genomic selection uses two types of datasets: a training set and a validation set. The training set is the reference population, in which the marker effects were estimated using certain statistical methods to incorporate with the phenotypic information and pedigree information. The validation set contains the selection candidates (derived from the reference population) that have been genotyped (but not phenotyped) and selected based on marker effects estimated in the training set. Implementation of genomic selection is likely to have major implications for genetic evaluation systems and for genetic improvement programmes generally (Goddard et al., 2007). In conclusion, the aim of genetic breeding is to reassemble desirable inherited traits to produce a species with improved characteristics, and each of the objectives can be addressed in a specific breeding technique.

2.3 Advance of fish disease resistance

Disease problems are a major cause of economic losses in fish aquaculture (Meyer, 1991). Considering the relationship between innate immunity and disease resistance, and based on experience from other farmed animals, selective breeding for resistant trait as a method of improving disease resistance seems attractive (Sweta and Sahoo, 2014); and it had attained successful application on different fish species, such as the selected lines against bacterial pathogens in rainbow trout and Atlantic salmon (Fevolden et al., 1992, 1993). The development of molecular-assisted techniques provides a chance to understand of the numbers of genes and their relative effects that determine expression of a trait. Although the genome-wide associated studies and genomic selection exhibit a promising application on selection breeding as talked above, they have not widely used in fish species because of the scarcity of genetic marker maps for most fish species. During the past years, QTL mapping methods is more often to detect loci associated with traits related to common stressors in commercial culture. So far, several QTL studies were conducted on varies diseases in cultured fish species. For instance, the infectious salmon anemia (ISA) in Atlantic salmon (Moen et al., 2004, Moen et al., 2007), the whirling disease in rainbow trout (Baerwald et al. 2011), the *Ceratomyxa shasta* resistance in salmonids (Nichols et al., 2003), the infectious hematopoietic necrosis virus (IHNV) in rainbow trout (Rodriguez et al., 2004), the infectious pancreatic necrosis in Atlantic salmon (Houston et al., 2008), the *A. salmonicida* resistance in turbot (Rodrıguez-Ramilo et al., 2011), the *Gyrodactylus salaris* resistance in Atlantic salmon (Gilbey et al., 2006), the Benedenia infections in

yellowtail (Ozaki et al., 2013). Besides, an allele of a microsatellite, Poli9-8TUF, associated with *lymphocystis* disease resistance (LD-R) in Japanese flounder was found to have a dominant effect at a single major locus (Fuji et al., 2006). A successful marker-assisted selection and further in farm trials confirm the feasibility of the selection process (Fuji et al., 2007). Various diseased associated SNPs in different fish species were also detected such as ISA in Atlantic salmon (Kjøglum et al. 2006), reovirus disease in grass carp (Wan et al., 2013). These promising results from selective breeding, along with the application of findings from molecular-assisted techniques, have shown that it is possible to develop highly resistant strains of fish.

3. Genomic and genetic techniques

3.1 Next generation sequencing techniques

Since first introduced to the market in 2005, the next-generation sequencing (NGS) technologies had a tremendous impact on genomic research. Therefore, several NGS platforms have been developed that provide low-cost, high-throughput sequencing. Of note were the 454 Genome Sequencers (Roche Applied Science), the Solexa Genome Analyzer (Illumina), the SOLiD platform (Applied Biosystems) and the HeliScope Single Molecule Sequencer (Helicos). The 454 system was the first next-generation sequencing platform available as a commercial product (Margulies et al., 2005). In 454 system, the sequencing is performed by the pyrosequencing method. Relative to other next-generation platforms, the key advantage of the 454 platform is read-length, which can reach to 500 bp per read currently. But the per-base cost of sequencing is much greater than that of other platforms. The Genome Analyzer (GA) was release by Solexa in 2006 and then Illumina purchased the company in 2007. The sequencer adopts the technology of sequencing by synthesis (SBS) (Fedurco et al., 2006). The Illumina genome analyzer is the most successful platform among next generation sequencing techniques because its output was 200 G per run initially, improved to 600 G per run currently which could be finished in 8 days. As of October 2014 all HiSeq instruments have been converted to the HiSeq 2500 platform, with each instrument capable of generating one terabase of sequence data per run. AB SOLiD platform has its origins in the system described by J.S. and colleagues in 2005 (Shendure et al., 2005) and then was purchased by Applied Biosystems in 2006. The sequencer adopts the technology of two-base sequencing based on ligation sequencing and consequently the sequence accuracy can reach to 99.85% after filtering. The Helicos single-molecule sequencing was developed in 2008 (Harris et al., 2008). It utilizes sequencing-by-synthesis methodology with a key advantage that the per-base raw substitution error rate (approaching 0.001%) may currently be the lowest of all the second-generation platforms. While the increasing usage and new modification in next generation sequencing, the third generation sequencing is coming out with new insight in the sequencing, such as PacBio sequencing and Nanopore sequencing. Relative to the second next-generation platforms, the third generation sequencing techniques have upgraded advantages including the shorten DNA preparation time for sequencing and signal capture in real time. In short, with the increasing maturity of the third generation sequencing technologies and market promotion,

genomics research is bound to revolutionize happen again.

3.2 BAC libraries construction

Large-insert genomic libraries are powerful tools for map-based gene cloning, comparative genomic studies and high-resolution physical map construction as well as the whole genome assembly (Smith et al., 1996). From the 1980s, large-insert genomic libraries, such as the Yeast Artificial Chromosome (YAC), the Bacterial Artificial Chromosome (BAC) and the P1-derived Artificial Chromosome (PAC) were developed and then widely used in genomics analysis. In particularly, BACs became widespread as an essential genomic resource because of the genetic stability, the lower chimeric rate and the higher conversion efficiency (Shizuya et al., 1992). For the construction of a BAC library, DNA is isolated and fragmented by hydroshearing, nebulization, or partial restriction digestion followed by preparative pulsed field gel fractionation. Fragments are size selected, ligated into a specialized vector, and maintained in *Escherichia coli* that are archived in 384-well plates at −80℃ (Chakrabarty et al., 2012). Where the BAC vector has a large fragment carrying capacity is the basic for BAC library construction. The quality of BAC libraries, typically represented by average insert size, and shorter DNA cannot fully utilize the advantage of BACs for the maximum cloning capacity. During the 1990s, a whole genome shotgun sequencing (WGS) approach was gradually adopted analysis of organisms having a large genome size (Adams et al., 2000), and paired-end sequencing of NGS is now drastically reducing the time and cost required for the WGS strategy. However, BAC library still play important roles in genome analysis, especially for analyzing genetically heterogeneous species and specific genomic regions with rich in repeat sequence and transposable elements. Therefore, several BAC libraries were constructed for aquatic animals, especially for fish species. For instance, the Atlantic salmon BAC library contains 305,557 clones with an average insert size of 188 kb, covering approximately 18 times the genome size (Thorsen et al., 2005); and the goldfish BAC library contains 128,352 clones with an average insert size of 150.4 kb, covering approximately 11 times the genome size (Luo et al., 2006). In addition, the other BAC libraries and associated resource are also available in zebrafish (http://www.sanger.ac.uk), medaka (Matsuda et al., 2001), tilapia (Katagiri et al., 2001), Japanese flounder (Katagiri et al., 2000), platyfish (Froschauer et al., 2002), channel catfish (Quiniou et al., 2003), and so on. Sequencing of BAC clones from both ends of inserts efficiently connects scaffolds separated by long gaps, which is difficult using paired-end sequencing of NGS. Besides, BAC sequences are also useful to assign sequence scaffolds onto chromosomes. In a word, in spite of the current widespread use of NGS, BAC libraries have irreplaceable roles in genome analysis as described above.

3.3 Restriction-site Associated DNA Sequencing

Genetic linkage map shows the relative locations of specific DNA markers along the chromosome, which was determined by the recombination rate during meiosis. It has the potential to facilitate the gene mapping, genome assembly and comparative genome studies (Jones et al., 2013). During the last few decades, the different types of molecular markers

were used in genetic map construction. These DNA markers are differentiates in two types. One is non-PCR based marker such as RFLP and another one is PCR-based marker such as RAPD, AFLP, SSR and SNP. Among these markers, the SSR has been the most widely used because of their polymorphism and transferability. To date, the SSR based genetic maps were constructed in a variety of fish species such as Atlantic salmon, Atlantic cod, sea bream, turbot, fugu, rainbow trout, tilapia, swordtail fish, medaka, three-spined stickleback, carp, channel catfish, guppies, European seabass, flounder and tongue sole (Yue, 2014). Although the SSR-based genetic map made a successful attempt to gene mapping and comparative analysis, the commonly number of markers, restricted by level of hundreds or few thousands, are inevitably difficult to carry out the fine-scale mapping and thus the much higher resolution of genetic map is being needed. The SNP is thus on the scene and has gained high attention because it represents the most abundant source of genetic variation in the genome. With the advances on next generation sequencing technology, the large-scale development of SNP markers and genotyping technologies have proliferated, with available platforms utilizing mass spectroscopy direct sequencing, fluorescence detection, and microchip hybridization. The technique of restriction-site associated DNA (RAD) tags, which provide a high-throughput manner and affordable cost of SNPs, has been used for SNP discovery and genotyping. It was particularly attractive for species with no reference genome (Etter and Johnson, 2012). The RAD-seq combines two simple molecular biology techniques with Illumina sequencing: the use of restriction enzymes to cut DNA into fragments, and the use of molecular identifiers (MID) to associate sequence reads to particular individuals. RAD-seq was first applied to investigation of the genetics of an important ecological trait in the three-spined stickleback, and then was gained wide applications not only in measurement of genetic diversity and investigation of population structure, but also in construction of high-resolution genetic maps (Baird et al., 2008). Kakioka et al. (2013) constructed the first genetic linkage map of *Gnathopogon* using a 198 F2 based on 1,622 RAD-tag markers. Genetic linkage maps also have been produced using RAD sequence data in stickleback (Baird et al., 2008), rainbow trout (Amish et al., 2012), guppies (Willing et al., 2011) and grouper (You et al., 2013). Besides, Hohenlohe et al. (2010) investigated whether evolutionary trajectories of stickleback were repeatable by comparing three independently derived freshwater populations and two oceanic populations from south Alaska using RAD-seq technique. Furthermore, the replicate parallel phenotypic evolution in stickleback may be occurring through extensive, parallel genetic evolution at a genome-wide scale, and the detection of population admixture and improved identification of hybrid and non-hybridized individuals using RAD-seq technique (Amish et al., 2012), suggest that the RAD-seq applicable utility in population genomics and phylogeography.

3.4 Whole Genome Bisulfite Sequencing

Epigenetics is defined as the study of heritable changes in gene expression without a change in DNA sequence (Bird et al., 2007). In general, epigenetic mechanisms include DNA methylation, chromatin structure and modification, and untranslated RNAs (Jones et al., 2002). Of them, DNA methylation is probably the most well researched epigenetic

marker that involves in multiple biological functions such as genomic imprinting, X-inactivation and tissue-specific gene expression (Jones et al., 2001). Several methods have been developed for detection of DNA methylation on a genome-wide scale such as Amplification of Inter-Methylated Sites (AIMS) method (Frigola et al., 2002), HpaII tiny fragments Enrichment by Ligation-mediated PCR (HELP) assay (Khulan et al., 2006) and the chromatin immune-precipitation of methylated fragments (MeDip) (Weber et al., 2005). Although these methods provide much epigenetic information on a species genome, they had limited coverage on CpG poor regions and in turn cannot provide a single-base resolution methylome. With the rapid development of next-generation sequencing technology drive the sequence price down to a point, a very promising method that is able to overcome limitations related to accurately quantify DNA methylation at the CpG resolution on a genome-wide level is Whole Genome Bisulfite Sequencing (WGBS). This method involves bisulfite conversion of unmethylated cytosines followed by adapter ligation and whole genome sequencing. It was originally tested on *Arabidopsis thaliana* genomic DNA because this species has a relatively small genome size (150 Mb), which made testing and troubleshooting cost efficient. From 55,805,931 aligned reads generated by WGBS, 39,113,599 were unique and nonclonal, yielding an average depth of 8.0 read coverage per base for each DNA strand, and overall unique read coverage of 78.5% of all cytosines in the *A. thaliana* genome with at least two reads (Lister et al., 2008). Since the first single-base resolution DNA methylomes for *A. thaliana* constructed, the other DNA methylomes have been sequenced such as human (Lister et al., 2009), silkworm (Xiang et al., 2010), mice (Stadler et al., 2011), corn (Gent et al., 2013), and soybean (Schmitz et al., 2013). However, few genome-wide methylation analyses of teleost species were reported. In fugu, the WGBS result showed that, as other vertebrates, the methylated cytosine sites in fugu are mainly distributed in CG context and further a relatively lower methylation level in the upstream of transcript start sites (TSS) was detected, indicating a negative correlation between methylation level and gene expression (Zemach et al., 2010). A more recent and high-throughput methylome of zebrafish was published. They generated nine single-base resolution DNA methylomes, including zebrafish gametes and early embryos. They found that the oocyte methylome is significantly hypomethylated compared to sperm and different inherent pattern between paternal and maternal DNA methylation during early embryogenesis, suggesting that the sperm DNA methylome is also inherited in zebrafish early embryos (Jiang et al., 2013). Although the WGBS is now relatively cheaper, it is still difficult to sequence large numbers of individual samples using this technique, especially for the species with bigger genome size. As a result, reduced representation bisulfite sequencing (RRBS) was developed in which 1% of the genome is studied making it possible to investigate large numbers of individuals at the cost of surveying the entire genome.

3.5 Whole genome sequence on teleost species

As early as 1920, Winkles cast term GENOME from GENes and chromosOMEs with the description of all genes and chromosomes in one species. In 1986, Thomas Roderick, an American scientist, puts forward a new concept-Genomics. It is a concept for all genes

which including genome mapping, nucleotide sequence analysis, gene localization and analysis of gene function (McKusick, 1997). Since the officially implementation of the Human Genome Project in the early 1990s, the genomics scientists were trying to sequence the genome completely of several organisms of important groups. The reason of sequencing the genome is it provides knowledge of total number of all genes; it shows relationship between genes; it provides opportunities to exploit the sequence for desired experimentation, all the genetic information of the organism and acts as an archive of all genetic information. The genomics revolution ushered in the rapid development of next-generation sequencing technology recently provides unprecedented opportunities for non-model species. In a word, the genomic era is now a reality.

For an ever-increasing number of teleost species a complete genome sequence is now known. The first fish with a completed genome sequence was the pufferfish. The compact genome of *Fugu rubripes* has been sequenced to over 95% coverage, and more than 80% of the assembly is in multigene-sized scaffolds. In this 365-megabase vertebrate genome, repetitive DNA accounts for less than one-sixth of the sequence, and gene loci occupy about one-third of the genome (Aparicio et al., 2002). Two years later, another pufferfish genome was also deciphered (Jaillon et al., 2004). The zebrafish genome-sequencing project was initiated in 2001. Now the Zv9 assembly version with a total size of 1.412 gigabases (Gb) provides evidence of more than 26,000 protein-coding genes (Howe et al., 2013). Another model medaka fish was also sequenced in 2007. The genome size of medaka is 700 Mb, which is less than half of the zebrafish genome. A total of 20,141 genes, including ~2,900 new genes, were annotated (Kasahara et al., 2007). Threespine sticklebacks is also a good model for studying the molecular basis of adaptive evolution in vertebrates and therefore its genome assembly was generated with a length-weighted median (N50) contig size of 83.2 kb, a length-weighted median (N50) scaffold size of 10.8 Mb and a total gapped size of 463 Mb. 20,787 protein-coding genes were annotated using the Ensembl pipeline (Jones et al., 2012). Besides, the first non-mode fish (Atlantic cod) genome was obtained exclusively by 454 sequencing of shotgun and paired-end libraries, and automated annotation identified 22,154 genes (Star et al., 2011). With the explosion of next generation sequencing, driving the sequence price down to a point, the non-model teleost species genome have made encouraging progress within recent years. In particularly a total of 5 teleost species genome assembly were obtained including sea lamprey (Smith et al., 2013), tuna (Nakamura et al., 2013), coelacanth (Amemiya et al., 2013), platyfish (Schartl et al., 2013) and zebrafish (Howe et al., 2013) in 2013. And to date in 2014, there are already 5 teleost species genome sequence were successful completed. All of them are considered as economical fish species including tongue sole (Chen et al., 2014), common carp (Xu et al., 2014), tilapia (Brawand et al., 2014), fugu (Gao et al., 2014), rainbow trout (Berthelot et al., 2014). Besides, at least 150 teleost species (most of them are economical fish) genome projects were initiated and some of them have been finished. In short, we can say that the development of the fish genome sequence has been the transition from model fish to economically important fish. I believe that in the near future, the vast majority of economical fish genome will be resolved. Such efforts will benefit to establish a platform on the level of genome-wide analysis of economically important traits and genome selection for aquaculture species.

4. Flatfish species and out targets

Flatfishes, members of the order Pleuronectiformes, comprise a biologically interesting and commercially important species for fisheries all over the world, with a unique asymmetric body shape developed as an adaptation to a lifestyle living on the sea bottom. This order comprises approximately 115 genera and around 700 extant species (i.e. hailihuts, flounders, soles, turbot, and plaice); the majority of them show a tropical distribution in the Indo-Pacific region (Sharina and Kartavtsev, 2010). With the general worldwide decline of wild fishery and an essentially stable wild catch of these fish species, investigation into producing them in aquaculture has been underway for the last thirty years. Aquaculture of some flatfish species like Japanese flounder, turbot, tongue sole, Atlantic halibut and others has now been successfully achieved, although improvements in the efficiency of production are still needed.

Half-smooth tongue sole (*Cynoglossus semilaevis*) is a commercially important cultured marine fish, distributing the coast of Temperate Zone in China. However, it is known that females grow faster and attain larger size than males. It thus deserves to produce monosex females' population in aquaculture using artificial sex control. The methods involved in the development of monosex require an understanding of the basis of sex determination. Previously study reveals that the tongue sole possesses a ZZ/ZW genetic sex determination system with a giant W chromosome. Interestingly, both wild and farmed tongue sole populations show systematically male-skewed sex ratio, suggesting a phenomenon of sex reversal in genetic female, although it was defined by exerting the genetic sex determination (GSD). Among the environmental factors implicated in sex differentiation, temperature is considered as a major determinant of reproductive activity and growth in fish (Chen et al., 2009). To date, the tongue sole have been subjected to intensive genetic studies, including the development of microsatellites (Liao et al., 2007) and female-specific DNA markers (chen et al., 2012), the characterization of sex-related genes (Dong et al., 2011), the artificial gynogenesis (Chen et al., 2009) and the construction of microsatellite linkage maps (Song et al., 2012). Therefore, the coexistence of GSD and ESD system within the tongue sole, coupled with the availability of its genomic and genetic information, provides an opportunity to comprehensively investigate the mechanism of sex determination and differentiation of this species.

Japanese flounder (*Paralichthys olivaceus*), one of the most desirable and highly priced marine fish species, is widely cultured along the coast of Northeast Asian countries such as China, Japan, and Korea. In addition to a booming Chinese market, more than 60,000 tons of Japanese flounder were artificially produced in Japan and Korea in 2013, making *P. olivaceus* one of the most important species in stock enhancement programs in these countries. With extensive cultivation, however, farming of the Japanese flounder has been confronted with certain problems, including a high mortality rate as well as a decline in growth. Recently, a number of genetic studies have been performed on the improvement of this species. For instances, at least four genetic maps have been constructed using various types of molecular markers (Coimbra et al., 2003; Kang et al., 2008; Castano-Sanchez et al., 2010; Song et al., 2012); and as followed, some economical traits were identified, such as a single major genetic locus associated with *lymphocystis* disease resistance (Fuji et al.,

2006) and its candidate gene TLR2 was identified (Hwang et al., 2011) as well as four QTLs for growth rate were identified on genetic linkage group (LG) 14 (Song et al., 2012). To further increase profitability and sustainability while maintaining genetic variability in the cultured stock, it is urgent to carry out both classic selective breeding and molecular marker assisted selection.

References:

1. Adams, M.D. et al. The genome sequence of *Drosophila melanogaster*. Science 287, 2185-2195 (2000).
2. Amemiya, C.T. et al. The African coelacanth genome provides insights into tetrapod evolution. Nature 496, 311-316 (2013).
3. Amish, S.J. et al. RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays. Molecular Ecology Resources 12, 653-660 (2012).
4. Aparicio, S. et al. Whole-genome shotgun assembly and analysis of the genome of *Fugu rubripes*. Science 297, 1301-1310 (2002).
5. Baerwald, M.R. et al. A major effect quantitative trait locus for whirling disease resistance identified in rainbow trout (*Oncorhynchus mykiss*). Heredity 106, 920-926 (2011).
6. Baird, N.A. et al. Rapid SNP discovery and genetic mapping using sequenced RAD markers. PloS One 3, e3376 (2008).
7. Baroiller, J.F., D'Cotta, H. and Saillant, E. Environmental effects on fish sex determination and differentiation. Sexual Development 3, 118-135 (2009).
8. Berthelot, C. et al. The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. Nature Communications 5, 3657 (2014).
9. Bird, A. Perceptions of epigenetics. Nature 447, 396-398 (2007).
10. Brawand, D. et al. The genomic substrate for adaptive radiation in African cichlid fish. Nature 513, 375-381 (2014).
11. Castano-Sanchez, C. et al. A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). BMC Genomics 11, 554 (2010).
12. Chakrabarty, S. et al. Comprehensive DNA copy number profile and BAC library construction of an Indian individual. Gene 500, 186-193 (2012).
13. Chen, S.L. et al. Artificial gynogenesis and sex determination in half-smooth tongue sole (*Cynoglossus semilaevis*). Marine Biotechnology 11, 243-51 (2009).
14. Chen, S.L. et al. Induction of mitogynogenetic diploids and identification of WW super-female using sex-specific SSR markers in half-smooth tongue sole (*Cynoglossus semilaevis*). Marine Biotechnology 14, 120-128 (2012).
15. Chen, S.L. et al. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. Nature Genetics 46, 253-260 (2014).
16. Cnaani, A. and Levavi-Sivan, B. Sexual development in fish, practical applications for aquaculture. Sexual Development 3, 164-75 (2009).
17. Coimbra, M.R.M. et al. A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. Aquaculture 220, 203-218 (2003).
18. Collard, B.C. and Mackill, D.J. Marker-assisted selection: an approach for precision plant breeding in the twenty-first century. Philosophical Transactions of the Royal Society B: Biological Sciences 363, 557-572 (2008).
19. Das, S. and Sahoo, P.K. Markers for selection of disease resistance in fish: a review.

Aquaculture International 22, 1793–1812 (2014).
20. Devlin, R. and Yoshitaka, N. Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208, 191-364 (2002).
21. Dong, X.L. et al. Molecular cloning, characterization and expression analysis of Sox9a and Foxl2 genes in half-smooth tongue sole (*Cynoglossus semilaevis*). Acta Oceanologica Sinica, 68-77 (2011).
22. Etter, P.D. and Johnson, E. RAD paired-end sequencing for local *de novo* assembly and SNP discovery in non-model organisms. Methods in Molecular Biology 888, 135-151 (2012).
23. Fedurco, M., Romieu, A., Williams, S., Lawrence, I. and Turcatti, G. BTA, a novel reagent for DNA attachment on glass and efficient generation of solid-phase amplified DNA colonies. Nucleic Acids Research 34, e22 (2006).
24. Fevolden, S.E., N.R., Refstie, T. and K.H. Roed. Disease resistance in Atlantic salmon (*Salmo salar*) selected for high or low responses to stress. Aquaculture 109, 215-224 (1993).
25. Fevolden, S.E., Refstie, T. and Røed, K.H. Disease resistance in rainbow trout (*Oncorhynchus mykiss*) selected for stress response. Aquaculture 104, 19-29 (1992).
26. Frigola, J., Ribas, M., Risques, R.A. and Peinado, M.A. Methylome profiling of cancer cells by amplification of inter-methylated sites (AIMS). Nucleic Acids Research 30, e28 (2002).
27. Froschauer, A. et al. Construction and initial analysis of bacterial artificial chromosome (BAC) contigs from the sex-determining region of the platyfish *Xiphophorus maculatus*. Gene 295, 247-54 (2002).
28. Fuji, K. et al. Marker-assisted breeding of a *lymphocystis* disease-resistant Japanese flounder (*Paralichthys olivaceus*). Aquaculture 272, 291-295 (2007).
29. Fuji, K. et al. Identification of a single major genetic locus controlling the resistance to *lymphocystis* disease in Japanese flounder (*Paralichthys olivaceus*). Aquaculture 254, 203-210 (2006).
30. Gao, Y. et al. Draft sequencing and analysis of the genome of pufferfish *Takifugu flavidus*. DNA Research 21(6): 627-637 (2014).
31. Genhua, Y. Recent advances of genome mapping and marker-assisted selection in aquaculture. Fish and Fisheries 15(3): 376-396 (2014).
32. Gent, J.I. et al. CHH islands: *de novo* DNA methylation in near-gene chromatin regulation in maize. Genome Research 23, 628-637 (2013).
33. Gilbey, J. et al. Identification of genetic markers associated with *Gyrodactylus salaris* resistance in Atlantic salmon *Salmo salar*. Diseases of Aquatic Organisms 71, 119-129 (2006).
34. Goddard, M.E. and Hayes, B.J. Genomic selection. Journal of Animal Breeding and Genetics 124, 323-330 (2007).
35. Harris, T.D. et al. Single-molecule DNA sequencing of a viral genome. Science 320, 106-109 (2008).
36. Hattori, R.S. et al. A Y-linked anti-Mullerian hormone duplication takes over a critical role in sex determination. Proceedings of the National Academy of Sciences of the

United States of America 109, 2955-2959 (2012).
37. Hohenlohe, P.A. et al. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. PloS Genetics 6, e1000862 (2010).
38. Houston, R.D. et al. A major QTL affects resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). Genetics, 1109-1115 (2008).
39. Howe, K. et al. The zebrafish reference genome sequence and its relationship to the human genome. Nature 496, 498-503 (2013).
40. Hwang, S.D. et al. Linkage mapping of toll-like receptors (TLRs) in Japanese flounder, *Paralichthys olivaceus*. Marine Biotechnology 13, 1086-1091 (2011).
41. Jaillon, O. et al. Genome duplication in the teleost fish *Tetraodon nigroviridis* reveals the early vertebrate proto-karyotype. Nature 431, 946-957 (2004).
42. Jiang, L. et al. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. Cell 4, 773-784 (2013).
43. Jones, D.B., Jerry, D.R., Khatkar, M.S., Raadsma, H.W. and Zenger, K.R. A high-density SNP genetic linkage map for the silver-lipped pearl oyster, *Pinctada maxima*: a valuable resource for gene localization and marker-assisted selection. BMC Genomics 14, 810 (2013).
44. Jones, F.C. et al. The genomic basis of adaptive evolution in threespine sticklebacks. Nature 484, 55-61 (2012).
45. Jones, P.A. and Baylin, S.B. The fundamental role of epigenetic events in cancer. Nature Review Genetics 3, 415-428 (2002).
46. Jones, P.A. and Takai, D. The role of DNA methylation in mammalian epigenetics. Science 293, 1068-1070 (2001).
47. Kakioka, R., Kokita, T., Kumada, H., Watanabe, K. and Okuda, N. A RAD-based linkage map and comparative genomics in the gudgeons (genus *Gnathopogon, Cyprinidae*). BMC Genomics 14, 32 (2013).
48. Kamiya, T. et al. A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). PloS Genetics 8, e1002798 (2012).
49. Kang, J.H., Kim, W.J. and Lee, W.J. Genetic linkage map of olive flounder, *Paralichthys olivaceus*. International Journal of Biological Sciences 4, 143-149 (2008).
50. Kasahara, M. et al. The medaka draft genome and insights into vertebrate genome evolution. Nature 447, 714-719 (2007).
51. Katagiri, T. et al. Construction and characterization of BAC libraries for three fish species; rainbow trout, carp and tilapia. Animal Genetics 32, 200-204 (2001).
52. Katagiri, T. et al. Genomic BAC library of the Japanese flounder *Paralichthys olivaceus.* Marine Biotechnology 2, 571-578 (2000).
53. Khulan, B. et al. Comparative isoschizomer profiling of cytosine methylation: the HELP assay. Genome Research 16, 1046-1055 (2006).
54. Kjøglum, S. et al. How specific MHC class I and class II combinations affect disease resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*). Fish and Shellfish Immunology 21, 431-441 (2006).
55. Knap P.W. and Bishop, S.C. Relationship between genetic change and in factious

disease in domestic livestock. Penicuik, UK: BSAS occasional publication, 65-80 (2000).
56. Lettre, G. and Rioux, J.D. Autoimmune diseases: insights from genome-wide association studies. Human Molecular Genetics 17, 116-121 (2008).
57. Lister, R. et al. Highly integrated single-base resolution maps of the epigenome in Arabidopsis. Cell 133, 523-36 (2008)
58. Lister, R. et al. Human DNA methylomes at base resolution show widespread epigenomic differences. Nature 462, 315-322 (2009).
59. Luo, J. et al. A BAC library for the goldfish Carassius auratus auratus (*Cyprinidae Cypriniformes*). Journal of experimental zoology. Part B, Molecular and developmental evolution 306, 567-574 (2006).
60. Margulies, M. et al. Genome sequencing in microfabricated high-density picolitre reactors. Nature 437, 376-380 (2005).
61. Marshall Graves, J.A. Weird animal genomes and the evolution of vertebrate sex and sex chromosomes. Annual Review of Genetics 42, 565-586 (2008).
62. Matsuda, M. et al. Construction of a BAC library derived from the inbred Hd-rR strain of the teleost fish, *Oryzias latipes*. Genes Genetics Systems 76, 61-63 (2001).
63. Matsuda, M. et al. DMY is a Y-specific DM-domain gene required for male development in the medaka fish. Nature 417, 559-563 (2002).
64. McKusick, V.A. Genomics: structural and functional studies of genomes. Genomics 45, 244-249 (1997).
65. Medugorac, I. et al. Conservation priorities of genetic diversity in domesticated metapopulations: a study in taurine cattle breeds. Ecology and Evolution 1, 408-420 (2011).
66. Meyer, F.P. Aquaculture disease and health management. Journal of Animal Science 69, 4201-4208 (1991).
67. Moen, T., Fjalestad, K.T., Munck, H. and Gomez-Raya, L. A multistage testing strategy for detection of quantitative trait Loci affecting disease resistance in Atlantic salmon. Genetics 167, 851-858 (2004).
68. Moen, T. et al. Mapping of a quantitative trait locus for resistance against infectious salmon anaemia in Atlantic salmon (*Salmo salar*): comparing survival analysis with analysis on affected/resistant data. BMC Genetics 8, 53 (2007).
69. Myosho, T. et al. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics 191, 163-170 (2012).
70. Nakamura, Y. et al. Evolutionary changes of multiple visual pigment genes in the complete genome of Pacific bluefin tuna. Proceedings of the National Academy of Sciences of the United States of America 110, 11061-11066 (2013).
71. Namekawa, S.H. and Lee, J.T. XY and ZW: is meiotic sex chromosome inactivation the rule in evolution? PloS Genetics 5, e1000493 (2009).
72. Nanda, I. et al. A duplicated copy of DMRT1 in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proceedings of the National Academy of Sciences of the United States of America 99, 11778-11783 (2002).
73. Navarro-Martin, L. et al. DNA methylation of the gonadal aromatase (cyp19a) promoter is involved in temperature-dependent sex ratio shifts in the European sea

bass. PloS Genetics 7, e1002447 (2011).
74. Nichols, K.M., Bartholomew, J. and Thorgaard, G.H. Mapping multiple genetic loci associated with *Ceratomyxa shasta* resistance in *Oncorhynchus mykiss*. Diseases of Aquatic Organisms 56, 145-154 (2003).
75. Ospina-Alvarez, N. and Piferrer, F. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. PloS One 3, e2837 (2008).
76. Ozaki, A. et al. Quantitative trait loci (QTL) associated with resistance to a monogenean parasite (*Benedenia seriolae*) in yellowtail (*Seriola quinqueradiata*) through genome wide analysis. PloS One 8, e64987 (2013).
77. Quiniou, S.M. et al. Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. Genetics Selection Evolution 35, 673-683 (2003).
78. Qvarnström, A. and Bailey, R. I. Speciation through evolution of sex-linked genes. Heredity 102, 4-15 (2009).
79. Rodriguez, M.F. et al. Genetic markers associated with resistance to infectious hematopoietic necrosis in rainbow and steelhead trout (*Oncorhynchus mykiss*) backcrosses. Aquaculture 241, 93–115 (2004).
80. Rodriguez-Ramilo, S.T. et al. QTL detection for *Aeromonas salmonicida* resistance related traits in turbot (*Scophthalmus maximus*). BMC Genomics 12, 541 (2011).
81. S. Jovanović, M.S. and D. Živković. Genetic variation in disease resistance among farm animals. Biotechnology in Animal Husbandry, 339-347 (2009).
82. Schartl, M. et al. The genome of the platyfish, *Xiphophorus maculatus*, provides insights into evolutionary adaptation and several complex traits. Nature Genetics 45, 567-572 (2013).
83. Schmitz, R.J. et al. Epigenome-wide inheritance of cytosine methylation variants in a recombinant inbred population. Genome Research 23, 1663-1674 (2013).
84. Shao, C.W. and Chen, S.L. Sex-determining genes and its association with mechanism of sex chromosome evolution in vertebrate. Journal of Agricultural Biotechnology 20, 1463-1474 (2012).
85. Sharina, S.N. and Kartavtsev Iu, F. Phylogenetic and taxonomic analysis of flatfish species (Teleostei, Pleuronectiformes) inferred from the primary nucleotide sequence of cytochrome oxidase 1 gene (Co-1). Genetika 46, 401-407 (2010).
86. Shendure, J. et al. Accurate multiplex polony sequencing of an evolved bacterial genome. Science 309, 1728-1732 (2005).
87. Shizuya, H. et al. Cloning and stable maintenance of 300-kilobase-pair fragments of human DNA in Escherichia coli using an F-factor-based vector. Proceedings of the National Academy of Sciences of the United States of America 89, 8794-8797 (1992).
88. Sinclair, A.H. et al. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346, 240-244 (1990).
89. Smith, C.A. et al. The avian Z-linked gene DMRT1 is required for male sex determination in the chicken. Nature 461, 267-271 (2009).
90. Smith, J.J. et al. Sequencing of the sea lamprey (*Petromyzon marinus*) genome provides insights into vertebrate evolution. Nature Genetics 45, 415-421 (2013).

91. Smith, T.P. et al. Construction and characterization of a large insert bovine YAC library with five-fold genomic coverage. Mammalian Genome 7, 155-156 (1996).
92. Song, W. et al. Construction of a high-density microsatellite genetic linkage map and mapping of sexual and growth-related traits in half-smooth tongue sole (*Cynoglossus semilaevis*). PloS One 7, e52097 (2012).
93. Song, W. et al. Construction of high-density genetic linkage maps and mapping of growth-related quantitative trail loci in the Japanese flounder (*Paralichthys olivaceus*). PloS One 7, e50404 (2012).
94. Stadler, M. B. et al. DNA-binding factors shape the mouse methylome at distal regulatory regions. Nature, 490-495 (2011).
95. Star, B. et al. The genome sequence of Atlantic cod reveals a unique immune system. Nature 477, 207-210 (2011).
96. Stelkens, R.B. and Wedekind, C. Environmental sex reversal, Trojan sex genes, and sex ratio adjustment: conditions and population consequences. Molecular Ecology 19, 627-646 (2010).
97. Thorsen, J. et al. A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6, 50 (2005).
98. Wan, Q. et al. Genomic sequence comparison, promoter activity, SNP detection of RIG-I gene and association with resistance/susceptibility to grass carp reovirus in grass carp (*Ctenopharyngodon idella*). Developmental Comparative Immunology 39, 333-342 (2013).
99. Waters, P.D., Wallis, M.C. and Marshall Graves, J.A. Mammalian sex--Origin and evolution of the Y chromosome and SRY. Seminars in Cell & Developmental Biology 18, 389-400 (2007).
100. Weber, M. et al. Chromosome-wide and promoter-specific analyses identify sites of differential DNA methylation in normal and transformed human cells. Nature Genetics 37, 853-862 (2005).
101. Wedekind, C. and Stelkens, R.B. Tackling the diversity of sex determination. Biology Letters 6, 7-9 (2010).
102. Willing, E.M. et al. Paired-end RAD-seq for de novo assembly and marker design without available reference. Bioinformatics 27, 2187-2193 (2011).
103. Wurschum, T. Mapping QTL for agronomic traits in breeding populations. Theoretical and Applied Genetics 125, 201-210 (2012).
104. Liao, X.L. et al. Polymorphic dinucleotide microsatellites in tongue sole (*Cynoglossus semilaevis*). Molecular Ecology Notes, 1147-1149 (2007).
105. Xiang, H. et al. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. Nature Biotechnology 28, 516-520 (2010).
106. Xu, P. et al. Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. Nature Genetics 46, 1212-1219 (2014).
107. Xu, Y.B. Molecular plant breeding. CABI (2012).
108. Yano, A. et al. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. Current Biology 22, 1423-1428 (2012).
109. Yoshimoto, S. et al. A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. Proceedings of the National Academy of

Sciences of the United States of America 105, 2469-2474 (2008).
110. You, X. et al. Construction of high-density genetic linkage maps for orange-spotted grouper *Epinephelus coioides* using multiplexed shotgun genotyping. BMC Genetics 14, 113 (2013).
111. Zemach, A. et al. Genome-wide evolutionary analysis of eukaryotic DNA methylation. Science 328, 916-919 (2010).

Chapter II BAC library construction and application in half-smooth tongue sole

Shao CW, Chen SL, Scheuring CF, Xu JY, Sha ZX, Dong XL and Zhang HB. Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. Marine Biotechnology, 2010, 12(5):558-568.

Shao CW, Liu G, Liu SS, Liu CL and Chen SL. Characterization of the cyp19a1a gene from a BAC sequence in half-smooth tongue sole (*Cynoglossus semilaevis*) and analysis of its conservation among teleosts. Acta Oceanologica Sinica, 2013, 32(8): 35-43.

Mar Biotechnol (2010) 12:558–568
DOI 10.1007/s10126-009-9242-x

ORIGINAL ARTICLE

Construction of Two BAC Libraries from Half-Smooth Tongue Sole *Cynoglossus semilaevis* and Identification of Clones Containing Candidate Sex-Determination Genes

Chang-Wei Shao · Song-Lin Chen · Chantel F. Scheuring · Jian-Yong Xu · Zhen-Xia Sha · Xiao-Li Dong · Hong-Bin Zhang

Received: 28 July 2009 / Accepted: 13 October 2009 / Published online: 3 December 2009
© Springer Science+Business Media, LLC 2009

Abstract Half-smooth tongue sole (*Cynoglossus semilaevis*) is an increasingly important aquaculture species in China. It is also a tractable model to study sex chromosome evolution and to further elucidate the mechanism of sex determination in teleosts. Two bacterial artificial chromosome (BAC) libraries for *C. semilaevis*, with large, high-quality inserts and deep coverage, were constructed in the *Bam*HI and *Hind*III sites of the vector pECBAC1. The two libraries contain a total of 55,296 BAC clones arrayed in 144 384-well microtiter plates and correspond to 13.36 haploid genome equivalents. The combined libraries have a greater than 99% probability of containing any single-copy sequence. Screening high-density arrays of the libraries with probes for female-specific markers and sex-related genes generated between 4–46 primary positive clones per probe. Thus, the two BAC libraries of *C. semilaevis* provided a readily useable platform for genomics research, illustrated by the isolation of sex determination gene(s).

Keywords Half-smooth tongue sole · BAC library · Sex determination · Female-specific marker · Sex-related gene · *Cynoglossus semilaevis*

C.-W. Shao · S.-L. Chen (✉) · J.-Y. Xu · Z.-X. Sha · X.-L. Dong
Key Laboratory for Sustainable Utilization of Marine Fisheries Resources, Ministry of Agriculture,
Yellow Sea Fisheries Research Institute,
Chinese Academy of Fishery Sciences,
106 Nanjing Road,
Qingdao, Shandong 266071, China
e-mail: chensl@ysfri.ac.cn

C.-W. Shao · C. F. Scheuring · H.-B. Zhang
Department of Soil and Crop Sciences and Institute for Plant Genomics and Biotechnology, Texas A&M University,
College Station, TX 77843, USA

Introduction

In contrast to mammals and birds, the sex-determining systems of fish appear to be at a primitive stage of evolution and display an amazing diversity, including male heterogametic, female heterogametic, and temperature-dependent systems (Matsuda 2003). The more complicated systems can be polyfactorial, involving multiple gene loci (Volff et al. 2007). Identification of the master sex-determining gene, which controls sex-determining mechanisms, should help to elucidate the whole sex determination cascade in teleosts. In mammals, the sex-determining gene, *SRY*, was identified on the human Y chromosome, but no orthologue of *SRY* has been found in other non-mammalian vertebrates (Sinclair et al. 1990). Recently, *DMY*, the first sex-determining gene among non-mammalian vertebrates was identified. *DMY* is at the top of the sex determination cascade in the teleost medaka, *Oryzias latipes* (Matsuda et al. 2002; Nanda et al. 2002). However, this gene, although sex-specific in medaka and *Oryzias curvinotus*, is not the master sex-determining gene in any other fish tested (Volff et al. 2003). More importantly, *SRY* and *DMY* are male sex-determining genes, carried on the Y chromosome of the male heterogametic (XX–XY) sex determination system (Sinclair et al. 1990; Matsuda et al. 2002). No female sex-determining gene in teleosts with the ZW heterogametic sex-determining system has so far been identified. Other sex-related genes involved in the sex-determining mechanism, such as Sox9 (Yokoi et al. 2002), *DMRT1* (Marchand et al. 2000), and P450arom (Kitano et al. 1999; Deng et al. 2009), are also important. Therefore, knowledge of the relationships between sex-determining genes and sex-related genes should help us to understand the fundamental pathways of fish sex determination and also to elucidate the

conserved mechanisms that operate behind sex determination in teleosts. Development of sex-specific molecular markers is an effective way to identify a sex-determining gene within a chromosome. Recently, male-specific DNA markers have been isolated and used for genetic sex identification in salmonids (Devlin et al. 2001; Devlin and Nagahama 2002). Also, sex-related amplified fragment length polymorphisms (AFLPs) and microsatellite markers were identified in the Nile Tilapia (*Oreochromis niloticus*; Lee et al. 2003; Ezaz et al. 2004). Two male-specific AFLP markers were also identified in the three-spined stickleback, *Gasterosteus aculeatus* L. (Griffiths et al. 2000). However, these species are of XY-type sex determination. Also, other model fish, such as green spotted pufferfish (*Tetraodon nigroviridis*) and zebrafish (*Danio rerio*), do not appear to be suitable for performing studies on genetic sex determination because of the absence of obvious sex-linked markers and heteromorphic sex chromosomes which can morphologically differentiated (Li et al. 2002; Schartl 2004).

Half-smooth tongue sole (*Cynoglossus semilaevis*), with its pleasant taste and high economic value, is an increasingly important marine fish species in China. In addition, it has been used as a model to study the sex determination mechanism. The half-smooth tongue sole has 20 pairs of euchromosomes and one pair of sex chromosomes (chromosomes Z and W) (Zhou et al. 2005). More importantly, the ZW sex-determining system of *C. semilaevis* is confirmed by the presence of WW individuals in artificial gynogenetic embryos (Chen et al. 2009). Seven female-specific AFLP markers have been isolated from the half-smooth tongue sole (Chen et al. 2007) and gynogenetic stocks of *C. semilaevis* have been developed (Chen et al. 2009). Thus, the half-smooth tongue sole not only has commercial value but is also an ideal model for the isolation of sex-determining gene(s) and for the elucidation of the mechanism of sex determination in teleosts. Positional cloning of sex-linked genes and genetic studies of *C. semilaevis* will be necessary to further understand the sex determination mechanism of this species and to develop molecular-assisted breeding programs. Recently, significant efforts have been made in half-smooth tongue sole genetics research to develop female-specific DNA markers (Chen et al. 2007), to characterize sex-related genes (Deng and Chen 2009), to develop molecular marker-assisted sex control (Chen et al. 2008), and to construct a genetic linkage map (Liao et al. 2009). In addition, a fosmid library of half-smooth tongue sole has been constructed with an average insert size of 39 kb, covering 3.23× haploid genome equivalents (Wang et al. 2009). However, this library may not be sufficient for positional cloning, genome physical map construction, and large-scale genome sequencing because of its relatively small insert size and low genome coverage.

Bacterial artificial chromosome (BAC) library are widely used in many areas of genomics and genetics research, including positional cloning (Zhang et al. 2008), integrative physical and genetic mapping (Xu et al. 2007), region-targeted marker development (Xu et al. 2006), and comparative genome analysis (Wang et al. 2007). Recently, the BAC libraries of some important fishes have been developed such as Japanese flounder (*Paralichthys olivaceus*), medaka (*O. latipes*), rainbow trout (*Oncorhynchus mykiss*), carp (*Cyprinus carpio*), tilapia (*O. niloticus*), platyfish (*Xiphophorus maculatus*), sea bream (*Pagrus major*), channel catfish (*Ictalurus punctatus*), cichlid (*Haplochromis chilotes*), Atlantic salmon (*Salmo salar*), European sea bass (*Dicentrarchus labrax*), threespine stickleback (*G. aculeatus*), swordtail fish (*Xiphophorus helleri*), and zebrafish (Takayuki et al. 2000; Matsuda et al. 2001; Katagiri et al. 2001; Katagiri et al. 2002; Froschauer et al. 2002; Quiniou et al. 2003; Masakatsu et al. 2003; Thorsen et al. 2005; Whitaker et al. 2006; http://bacpac.chori.org/home.htm). However, no BAC libraries have been previously constructed for half-smooth tongue sole. Thus, we have constructed two BAC genomic libraries from the genomic DNA of female half-smooth tongue sole. BAC clones which contained a candidate master sex-determining gene or sex-related genes were identified by hybridization screening. These two BAC libraries provide sufficient clone resources for comprehensive genome research, physical map development, and positional gene cloning in *C. semilaevis*. More importantly, in comparison to the other fish species that have been used to study mechanisms of sex determination, the half-smooth tongue sole has well-defined sex chromosomes, with sex-linked molecular markers allowing the positional cloning of the master sex-determining gene. Thus, half-smooth tongue sole will probably contribute considerably to our understanding of the molecular and evolutionary mechanisms driving the variability of sex determination in fish.

Results and Discussion

Sex Identification

Three fish were initially determined as females based on morphological examination. Following sacrifice, routine histology also showed that the gonads of the three selected fish were ovaries containing of oocytes and not testis containing spermatocytes (Fig. 1). However, the approaches used above only identify the phenotypic sex in *C. semilaevis*. Since *C. semilaevis* can undergo the natural phenomenon of sex-reversal, it was necessary to use a female-specific AFLP marker (CseF783) to determine the genetic sex of the selected fish (Chen et al. 2007). The three

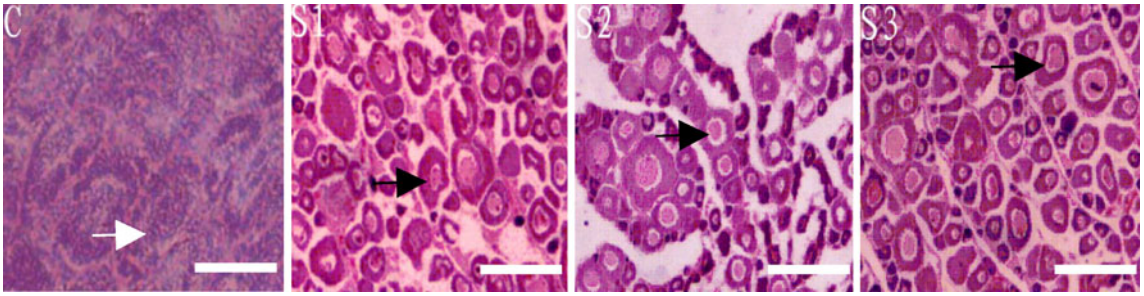

Fig. 1 Routine histology of the gonads. *C* is the control male testis and *S1–S3* are the ovaries of the three female samples. *White arrowhead* indicates the spermatocytes and *black arrowheads* indicate the oocytes. Bar=5 μm

sample fish, identified morphologically as female, along with a female control produced a 749 bp band when screened with the female-specific marker, as shown in Fig. 2. This band was absent from the male control fish. The three individuals that were phenotypically and genetically identified as females were used for BAC library construction.

Genome Size Estimation

To calculate the coverage of the BAC library, the genome size of the half-smooth tongue sole was estimated by flow cytometry. Flow cytometry is a well-established method for nuclear DNA content analysis and characterization (Peruzzi et al. 2005). The DNA content (C value) defined as the amount of DNA per haploid nucleus is a fundamental parameter of genomics studies. The C values of the three half-smooth tongue sole individuals were 0.6462, 0.6541, and 0.6228 pg, respectively (Fig. 3). The mean nuclear DNA content (C value) of female *C. semilaevis* was 0.6410 pg. Using the conversion factor of 1 pg=978 Mb (Dolezel et al. 2003), the genome size of the half-smooth tongue sole is estimated to be 626.9 Mb. This is consistent with a previous report of 606.4 Mb (Wang et al. 2009). In comparison to the genome sizes of the other 33 flatfish species (0.39–1.10 pg) (Wang et al. 2009), the half-smooth tongue sole has a relatively small genome size which is advantageous for genomic studies.

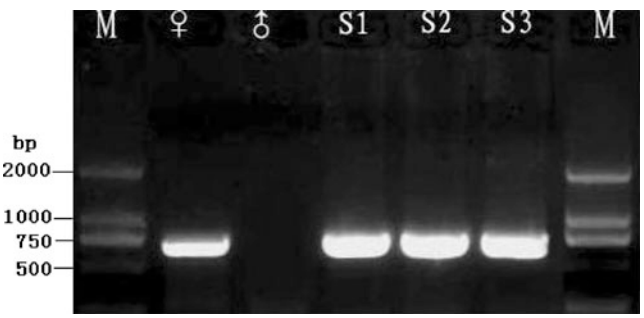

Fig. 2 Genetic sex determination of half-smooth tongue sole fish using a female-specific marker (CseF783). Female (♀) and male (♂) controls are indicated. *S1–S3* are experimental samples. *M* is DNA ladder DL2000 (TaKaRa)

BAC Library Construction and Characterization

Two BAC libraries were constructed from the nuclear DNA of female *C. semilaevis* in the BAC vector pECBAC1. The library constructed by *Bam*HI partial digestion consists of 15,360 clones. The other library, constructed by *Hind*III partial digestion, consists of 39,936 clones (Table 1). To evaluate the quality of the BAC libraries, 100 BAC clones were randomly picked from each of the two libraries, and their plasmids were isolated. After digestion with *Not*I and separation of the fragments by pulsed-field gel electrophoresis (PFGE), it was possible to resolve the vector DNA band (7.5 kb) and one or more bands of insert DNA, depending on the number of *Not*I sites in the cloned insert. The size of each clone was estimated manually using a standard pulsed-field DNA size marker (New England Biolabs, Low Range Marker) as a reference. The average insert size was 160.0 kb for the *Bam*HI library, with inserts ranging from 60 to 295 kb and 155.0 kb for the *Hind*III library, with inserts ranging from 45 to 280 kb (Fig. 4). The very large average insert size increases the probability of obtaining complete gene sequences within a single BAC clone and will be beneficial for physical mapping (Thorsen et al. 2005). Ninety-nine *Bam*HI and 96 *Hind*III BACs contained inserts and only one *Bam*HI (1%) and four *Hind*III (4%) BACs had no inserts. The low frequency of non-insert clones will increase the efficiency of transformation and reduce the costs of library storage and utilization. The insert size distribution fits well with the size of the DNA fragments that were used in the cloning procedure. The majority of DNA inserts fell into the 120 to180 kb range, and only 5% of DNA inserts were less than 100 kb (Fig. 5). The two BAC libraries consisted of 55,296 clones with an average insert size of 156.4 kb, which were manually arrayed into 144 384-well plates. Based on a genome size of 626.9 Mb, the two libraries together provide 13.36 haploid genome equivalents, and they will enable the isolation of clones containing virtually any gene of interest.

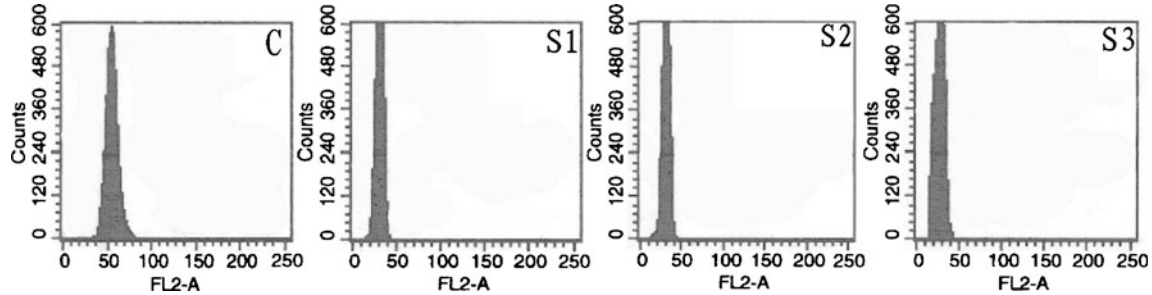

Fig. 3 Flow cytometric analysis of DNA content of *C. semilaevis* erythrocytes (*S1–S3*) and standard chicken erythrocytes (*C*). DNA values are reported in arbitrary units expressed as fluorescence channel numbers (FL2-area). The channel number of the control is 58.57. The channel numbers of the samples *S1–S3* are 30.28, 30.65, and 29.18, respectively. Thus, the DNA content (*C* value) of each sample is 0.6462, 0.6541, and 0.6228 pg, respectively, following the formula C=(CnS/CnC)×1.25. (CnS is the channel number of sample; CnC is the channel number of the control; the C value of chicken erythrocytes is 1.25)

Also, we tested the stability of the BAC clones by continuously growing eight random clones for 5 days, followed by analyzing the restriction patterns of the DNA of the culture cells on a pulsed-field agarose gel (Fig. 6). Identical restriction patterns were observed between the DNAs of all eight clones isolated from the day 1 and day 5 cultures, indicating that the clones of the libraries are stable in the bacterial host for at least 100 generations. Moreover, since the two libraries were developed using two different restriction enzymes with non-complementary recognition sites (*Hind*III is AT-rich and *Bam*HI is GC-rich), it is expected that the genome coverage of these combined libraries will be more representative of the half-smooth tongue sole genome than the coverage based on a library constructed with a single enzyme. This is especially important for the development of physical maps using the libraries (Chang et al. 2001); in the near future, map-based cloning may become a reality in *C. semilaevis*.

BAC Library Screening

To isolate the master sex-determining gene and other sex-related genes, we screened the two BAC libraries with ten overgo (overlapping oligonucleotides) probes (Table 2). Five of the ten probes (O-cyse01-05) are female-specific and were designed from genomic DNA sequence; the other probes (O-cyse06-10) were designed from either 3′ UTR or coding sequences of the sex-related genes. High-density filters representing the two half-smooth tongue sole BAC libraries were screened with two sets of five-pooled probes, resulting in the identification of 454 putative positive clones. This demonstrated the feasibility of our screening strategy using overgo probes and hybridization (Fig. 7). To identify the correspondence between clones and individual probes, a second hybridization experiment was performed for each probe individually with new nylon filters containing all 454 clones. Of the 454 BAC clones identified in the primary screen, 226 were verified in these secondary screening experiments. The remaining 228 clones were not positive clones because weak signals were also identified as positive clones in the first hybridization.

Each sex-related gene probe identified an average of 15.2 positive clones (ranging from four to 25), with a total of 76 positive BAC clones for the five gene probes. Taking into consideration empty insert rate, the mean number of positive clones per sex-related gene probe is consistent with the library coverage (13.36×genome equivalents; Table 3). However, an average of 30 positive clones were identified for each female-specific probe (ranging from 15 to 46), which is almost two times higher than expected for single-copy genes (Table 3). This can be explained if the half-smooth tongue sole genome contains multiple copies of the female-specific sequences. However, this observation could also result from uneven distribution of restriction enzyme sites along the genome and/or unbalanced coverage of the two libraries. Further screening of the libraries with a large

Table 1 Summary of the two-tongue sole BAC libraries

Library name	Clone vector	Restriction enzyme used	Total plates	Total clones	Insert-empty clones (%)	Average insert size (kb)	Genome coverage
*Bam*HI library	pECBAC1	*Bam*HI	40	15,360	<1	160	3.88×
*Hind*III library	pECBAC1	*Hind*III	104	39,936	<4	155	9.48×
Combined library			144	55,296	<3.2	156.4	13.36×

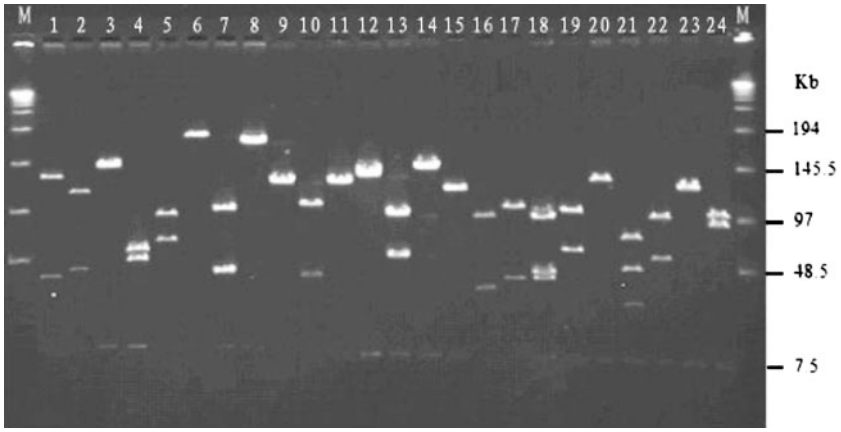

Fig. 4 Insert size analysis of the *Hind*III *C. semilaevis* BAC library. DNA was extracted from randomly selected clones and digested with *Not*I to release the vector (the 7.5 kb band). The *first* and *last lanes* (*M*) are mid-range II PFG markers and the *lanes* between the markers are the *Not*I digested BAC DNAs

number of single-copy sequences will be needed to accurately estimate the overall genome coverage of the libraries (Tao et al. 2001). Furthermore, a possible over-estimation of the genome size and an under-estimation of the insert size may also partly account for the lower genome coverage estimated by screening with the overgo probes.

We found that seven true-positive clones [Bam017M07 (CseF382 and CseF783), Bam021M07 (CseF382 and CseF464), Bam024M04 (CseF382 and CseF783), Bam010B01 (CseF305 and CseF464), Bam038J11 (CseF136 and CseF464), Hind012D10 (CseF382 and CseF136), and Hind027L15 (CseF382 and CseF464)] contained two female-specific sequences. Two sex-related genes (P450B and Dmrt1a) were identified in one positive clone (Hind015C15). This further explained the imbalance between the coverage and the number of positive clones. The location of more than one gene of interest will also make physical mapping feasible. Further analysis of the true-positive clones that contain female-specific sequences and sex-related genes will help us to identify the female sex-determining gene and to understand the mechanism of sex determination in *C. semilaevis*.

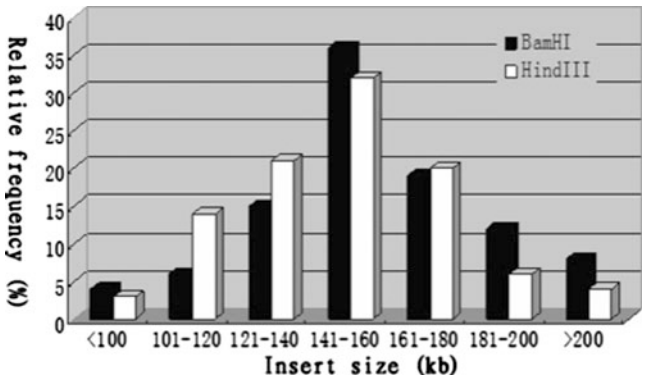

Fig. 5 Insert size distribution in the two *C. semilaevis* BAC libraries. The black histogram (*filled square*) represents the *Bam*HI library, which has an average insert size of 160 kb. The white histogram (*open square*) represents the *Hind*III library, which has an average insert size of 155 kb. The average insert size of the combined library was 156.4 kb

Materials and Methods

Fish and Sample Collection

Three half-smooth tongue sole individuals with an average weight of 2–2.5 kg were collected from a commercial hatchery (Laizhou Mingbo Aquatic Co., Laizhou City, China). Blood was taken from the individuals using sterile syringes, which had been pre-prepared with anticoagulant solution (0.5 M EDTA, pH 8.0). Blood cells were stored at

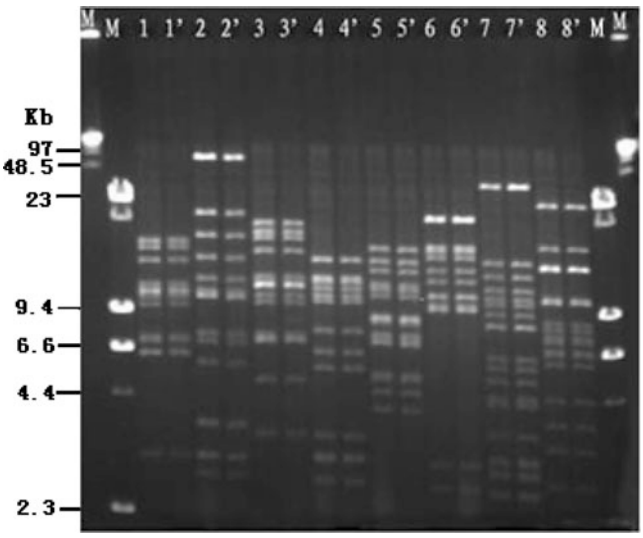

Fig. 6 Stability detection of the half-smooth tongue sole BAC clones. Clones having insert size of 150 kb or larger were selected, grown continuously for 5 days, and analyzed by restriction digestion with *Hind*III. *M* lambda ladder PFG marker, *1–8* restricted patterns of the DNA isolated from day 1 cells; and 1′–8′: restricted patterns of the DNA isolated from day 5 cells

Table 2 Description of the ten overgo probes from sex-specific AFLP markers or sex-related genes

Probe designation	GenBank accession number	Sex-specific AFLP markers and sex-related genes	Input sequence region for overgo design	GC content (%)/Tm (°C)	Left (L)and right (R) primer sequence
O-cyse01	DQ487760	CseF382	Genomic DNA	52.6%	L:AATTCACTGACCCCTGAGAGCGG
				68.7°C	R:AACTGCCTAACTTCGCCGCTCTC
O-cyse02	EU430410	CseF136	Genomic DNA	55.3%	L:GGGTGGAGACCTAGTTAGTGCCT
				68.4°C	R:AAAAGCCCGGGTAGAAGGCACTA
O-cyse03	EU430411	CseF305	Genomic DNA	57.9%	L:GACTGCGTACCAATTCACCACCTTC
				69.2°C	R:GATTTTTTCCTTCTCTGGAAGGTGG
O-cyse04	EU430412	CseF464	Genomic DNA	52.6%	L:GCTGCACCTCTAGACTTATCCCA
				67.4°C	R:TGGCAGGGAAAAGAGTGGGATAA
O-cyse05	EU430413	CseF783	Genomic DNA	48.7%	L:GTTTAGAGTGCGGACCGTAGAC
				65.1°C	R:GAAGTGTAGTTGGCATGTCTACGG
O-cyse06	EU070762	Sox9	Coding sequence	52.6%	L:TACGCGTGAACGGATCTACGAAG
				67.2°C	R:GACGTGAGGCTTGTTCTTCGTAG
O-cyse07	EU070763	Sox10	3′ UTR	50.0%	L:ACCAGAACAAGACTGCAGCGAAG
				67.7°C	R:TGGTTGACACACTGACTTCGCTG
O-cyse08	EF198239	P450B	3′ UTR	46.3%	L:CATGTTGAAGTGTCTGGACAACATC
				66.4°C	R:CAATTGACCCATGTTCCGATGTTGT
O-cyse09	EF134716	P450A	3′ UTR	46.3%	L:AGGTAATTGGCAGCACGCTGTTGA
				66.5°C	R:GGTTTGTAGCGTCATTATCAACAGC
O-cyse10	EU070761	Dmrt1a	3′ UTR	45.2%	L:CTAGTCTGGATGTTGTTCATAGTCA
				64.3°C	R:GGCTGGGGTGAGAATATTGACTATG

4°C for use in the estimation of genome size and for the preparation of high-molecular-weight DNA. Muscle tissue of the three fish was dissected and stored at −80°C for DNA isolation. The preserved DNA of the control male and female which had been identified by histological gonad analysis were stored at −80°C in our laboratory.

Sexing of Fish by Histological Gonad Analysis and Female-Specific AFLP Marker Typing

All fish were first sexed by macroscopic examination of the gonads. Gonads were fixed in Bouin's fixative for 12 h and were sexed using routine histological procedures, including hematoxylin–eosin staining. To further confirm the sex, the DNAs of the three fish were extracted from muscle tissue using a standard proteinase K digestion, phenol–chloroform extraction, and RNase treatment protocol (Liao et al. 2007; Shao et al. 2008). Genetic sex identification was performed using a female-specific AFLP marker, as previously described (Chen et al. 2007). A pair of female-specific polymerase chain reaction (PCR) primers (CseF783F: 5′ TGATGACACAGGATACGTCTGC3′ and CseF783R: 5′ TCTGTGGCACTTAGTGGGATGT3′) was used for PCR. A female-specific DNA band of 749 bp was amplified from

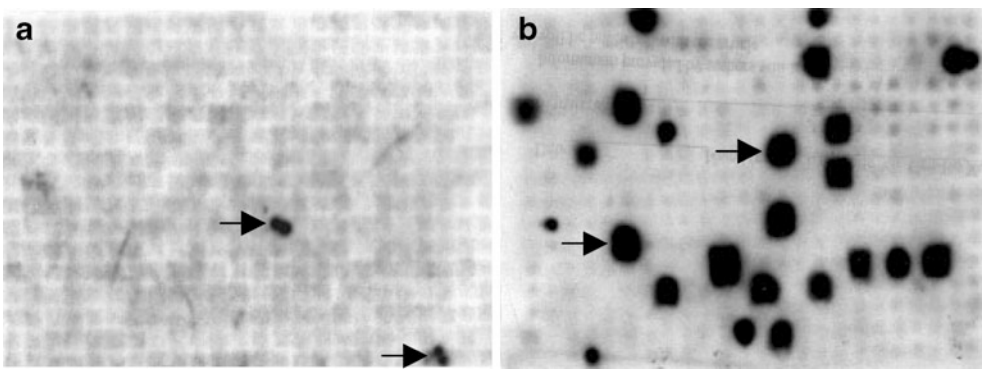

Fig. 7 Example of positive clones identified from the libraries using overgo probes. **a** Positive clones are indicated by double hybridization signals because each clone was double-spotted on the filter in first hybridization. **b** One-spotted positive clones in second hybridization. *Arrows* showed the double-spotted and one-spotted positive clones

Table 3 The results of the tongue sole BAC libraries screening with sex-related gene probes and sex-specific AFLP marker probes

Sex-related genes and sex-specific AFLP markers	Number of hits	Clone positions in the library
Sox9	18	Bam008L01, Bam010I18, Bam017P21 Hind002I02, Hind015G24, Hind019F14, Hind022H05, Hind023B04, Hind049N05 Hind062E01, Hind066C11, Hind066P10, Hind069H02, Hind081K03, Hind094D01 Hind094P14, Hind096A06, Hind102A07
Sox10	20	Bam001E21, Bam010M09, Bam015L02, Bam026J06, Bam027B07, Bam029D05 Bam030K18, Bam034G03, Bam037E24 Hind007N14, Hind008P03, Hind009C06, Hind013E24, Hind021F08, Hind033H17 Hind047P03, Hind049A09, Hind075M09, Hind077K05, Hind099M01
P450B	25	Bam007C16, Bam014L03, Bam021A04, Bam039D20 Hind001N04, Hind005D07, Hind007I03, Hind012E04, Hind012G22, Hind015C15[a] Hind016N14, Hind027A23, Hind031D16, Hind035G19, Hind042K08, Hind050O07 Hind051G19, Hind054P07 Hind056H03, Hind062O21, Hind063L03, Hind066D13, Hind067G04, Hind070C16 Hind085B23
P450A	4	Bam002D14, Bam023G20 Hind025N15, Hind085M14
Dmrt1a	9	Bam015H22, Bam018D20 Hind003O24, Hind004B04, Hind015C15[a], Hind016K22, Hind076G15, Hind079P02 Hind82P06
CseF382	27	Bam010C18, Bam013H14, Bam017M07[b], Bam021M07[b], Bam021G14 [b]Bam024M04, Bam029B14, Bam030C23 Hind005A11, Hind011L12,Hind012D10[b], Hind017H20, Hind021I02, Hind022E18 Hind025J17, Hind027L15[b], Hind036K02, Hind037D16, Hind059D22, Hind059G18 Hind060C14, Hind066J22, Hind082K04, Hind089C15, Hind093J11, Hind098D08 Hind100K18
CseF464	38	Bam002C20, Bam010B01[b], Bam010D19, Bam019L05, Bam020N17, Bam021M07[b] Bam025I03, Bam027I19, Bam028A02, Bam028D22, Bam030H17, Bam035P03 Bam038J11[b] Hind010N04, Hind012E04, Hind014B05, Hind017N07, Hind020M14, Hind022G03 Hind027L15[b], Hind034C07, Hind036D22, Hind036G01, Hind042H20, Hind043N14 Hind046E14, Hind054N21, Hind064F06, Hind074B16, Hind078P13, Hind081J24 Hind082H18, Hind084F21, Hind087O14, Hind088E03, Hind094K01
CseF783	46	Bam006M12, Bam009P14, Bam011B08, Bam017M07[b], Bam024M04[b], Bam032B07 Bam034D12, Bam038I06 Hind001F06, Hind004K09, Hind010C04, Hind014K24, Hind022J15, Hind024C04 Hind025O13, Hind028H15, Hind029G14, Hind031L03, Hind032B19, Hind032J20 Hind034C02, Hind034N18, Hind035B10, Hind036D15, Hind036J04, Hind039E13 Hind041K19, Hind048H09, Hind048E04, Hind055F08, Hind056M14, Hind057J06 Hind059I20, Hind061E24, Hind061P12, Hind069A02, Hind074C23, Hind074P07 Hind078J07, Hind079A17, Hind088J07, Hind089B18, Hind091F23, Hind094G15 Hind097F07, Hind099M06, Hind102C15
CseF305	24	Bam006M04, Bam007H11, Bam008B21, Bam010B01[b], Bam016N21, Bam032P24 Hind002G19, Hind007E07, Hind012B21, Hind012J22, Hind029O14, Hind036J04 Hind044P04, Hind047G17, Hind048C16, Hind050C04, Hind063H07, Hind063F11 Hind066F11, Hind069C07, Hind076N11, Hind082E09, Hind084C23, Hind095H15
CseF136	15	Bam001M20, Bam008L12, Bam016D04, Bam019H09, Bam037O06, Bam038J11[b] Hind005G01, Hind012D10[b], Hind018A20, Hind022G03, Hind025D04, Hind026P03 Hind060C21, Hind092B12, Hind104P14[b]

[a] Indicated that two sex-related genes (P450B and Dmrt1a) were identified in one positive clone (Hind015C15)

[b] Indicated that seven true-positive clones [Bam017M07 (CseF382 and CseF783), Bam021M07 (CseF382 and CseF464), Bam024M04 (CseF382 and CseF783), Bam010B01 (CseF305 and CseF464), Bam038J11 (CseF136 and CseF464), Hind012D10 (CseF382 and CseF136), and Hind027L15 (CseF382 and CseF464)] contained two female-specific sequences

DNA of female individuals while no 749 bp band was amplified from DNA of male individuals.

Estimation of the Half-Smooth Tongue Sole Genome Size

Blood samples were prepared for propidium iodide flow cytometric analysis as described by Tiersch et al. (1989). Fresh chicken blood was used as an external reference. The blood cells were washed three times and then diluted to 10^5 cells per milliliter using phosphate buffered saline (PBS) (0.8% NaCl, 0.02% KCl, 0.144% Na_2HPO_4, and 0.024% KH_2PO_4, pH 7.4). The samples were suspended in 0.5 ml of 50 mg/L propidium iodide solution, containing 25 μl RNAase (1.0 mg/ml), 0.1% sodium citrate, and 0.1% Triton X 100. After mixing, the samples were filtered through a 20-μm nylon mesh and analyzed within 15 min by flow cytometry. The mean DNA content of at least 10^3 cells per sample was measured with a FACScan Flow Cytometer (Becton Dickinson) and is reported in arbitrary units expressed as fluorescence channel numbers (FL2-area).

Preparation of High-Molecular-Weight DNA

High-molecular-weight DNA was prepared according to Zhang et al. (1995), with minor modifications (Zhang et al. 2009). Five milliliters of the stored blood was transferred to two 50 ml conical screw-cap polypropylene tubes and three volumes of ice-cold PBS were added to each tube and mixed gently, followed by centrifugation at 1,500 g for 20 min. The cell pellet was washed an additional three times with ice-cold PBS under the same conditions. The cell pellet was resuspended and diluted to 1×10^8 cells/ml with PBS. The cell suspension was immediately mixed with an equal volume of pre-warmed (50°C) 1% low melting point (LMP) agarose in PBS. The cell/molten agarose mixture was pipetted into100-μl plug molds (Biorad), and plug molds were incubated at 4°C for 15 min to solidify the agarose. The final agarose concentration was 0.5%, and the cell concentration was 5×10^7/ml. The cells embedded in the LMP agarose plugs were lysed by incubation in lysis buffer (0.5 M EDTA; pH 9.0–9.3, 1% sodium lauryl sarcosine, and 0.3 mg/ml proteinase K) for 24 h at 50°C with gentle rotation. Plugs were washed three times in Tris-EDTA (TE) buffer (10 mM Tris HCl, 1 mM EDTA, pH 8.0) containing 40 mg/ml phenylmethylsulfonyl fluoride (PMSF) for 30 min at room temperature, and then the plugs were washed three times in ice-cold TE buffer for 30 min to remove PMSF. Plugs were stored in 0.5 M EDTA (pH 9.0–9.3) at 4°C.

Construction of the Half-smooth Tongue Sole BAC Library

The construction of the BAC library was performed as described by Zhang and Wu (2001) and Ren et al. (2005). Briefly, prior to partial digestion with *Bam*HI or *Hind*III, the DNA plugs were washed three times (1 h each time) in 10–20 volumes of ice-cold TE buffer. Each plug was cut into nine smaller pieces approximately equal in size, and these were incubated in incubation buffer [1× React buffer (Gibco BRL, Grand Island, NY) and 2 mM spermidine, 1 mM DTT (dithiothreitol), and 0.2 mg/ml bovine serum albumin] on ice for 1 h. Three pieces of the plug were transferred into 170 μl digestion buffer with various concentrations of *Bam*HI or *Hind*III, and reactions were incubated on ice for an additional 1.5 h. The reactions were transferred to a 37°C water bath, incubated for exactly 8 min, then a 1/10 volume of 0.5 M EDTA pH 8.0 was added to the tubes to stop the reactions. The partially digested samples were resolved on 1% agarose gels run in 0.5×tris borate EDTA buffer (TBE) buffer by PFGE at 6 V/cm, 12.5°C with 50 s switch times, and a linear ramp, for 18 h. Optimum enzyme concentration was determined by visualizing maximum fragment concentration in the 100 to 400 kb range. Based on these pre-determined enzyme concentrations, eight plugs were treated as described above for large-scale partial digestions. Digestions were separated on a 1% LMP agarose gels in 0.5× TBE buffer with the following PFGE conditions, 6 V/cm, 12.5°C with a 90 s switch time for 9 h, followed by 4 V/cm, and 12.5°C with a 5 s switch time for 7 h. Markers on the gel were stained for visualization. The gel region containing 100 to 400 kb DNA fragments was collected, and the DNA fragments were recovered from the agarose gel slices by electroelution in dialysis tubing (Spectra/Pro Dialysis 7, Spectrum Laboratories Inc.) at 6 V/cm, 12.5°C, and a 35 s switch time in 0.5× TBE for 4 h, followed by reversing the polarity of the current for 60 s. The DNA was gently collected with a wide-bore pipette tip, and then a second round of size selection was applied to the collected DNA fragments to further purify the desired size of DNA fragments and to purge small trapped fragments. Electrophoresis parameters were set to 4 V/cm, 12.5°C, and a 5 s switch time for 8 h. In the second size selection gel, the gel containing the size fraction larger than 150 kb was excised and recovered by electroelution, as above. The eluted DNA fragments were dialyzed in the same tubing against 0.5× TE for at least 3 h at 4°C with a change of TE every hour.

The concentration of the insert DNA was checked on a conventional 1% agarose gel using known concentrations of lambda DNA as a standard. DNA fractions were ligated into the *Bam*HI or *Hind*III sites of pECBAC1 at 16°C overnight in a total volume of 100 μl with a vector: insert ratio of 5:1. Then 1.5 μl of the ligated DNA was electroporated into 20 μL of *Escherichia coli* DH10B competent cells. The electroporation procedure was carried out on a Cell Porator and Voltage Booster System (Gibco BRL) at 350 V, 330 μF capacitance, low ohm impedance, and fast charge rate, and

the Voltage Booster was set at 4,000 Ω. Electroporated cells were immediately diluted in 1 ml of SOC medium and incubated at 37°C for 1 h with shaking at 200 rpm and then plated on Luria-Bertani (LB) agar medium containing 12.5 μg/μl chloramphenicol, 0.55 mM IPTG, and 80 μg/ml X-Gal, followed by incubation at 37°C for 24 h.

The average insert size of white clones from each ligation was determined by analyzing 50 to 100 random BAC clones. The ligations that gave 600 white clones per ligation transformation with the largest insert sizes were chosen for the construction of the libraries and were transformed into DH10B competent cells as above. White colonies were picked into 384-well microtiter plates containing 65 μl of freezing medium in each well (Zhang et al. 1996). After incubation overnight at 37°C, the microtiter plates were frozen and stored at −80°C. From the original copy of the BAC library, two more copies were reproduced and stored separately.

Characterization of the Half-Smooth Tongue Sole BAC Library

To check the insert sizes of the clones, randomly selected white colonies were inoculated into 5 ml LB broth containing chloramphenicol (12.5 μg/ml) and grown overnight at 37°C with agitation at 250 rpm. Cells were harvested, and DNA was isolated using the alkaline lysis method, as described by Zhang and Wu (2001). For insert size estimation, 8 μl of BAC DNA was incubated at 37°C for 3 h with 0.2 U of *Not*I enzymes in a 40-μl reaction to release the insert DNA from the vector. The *Not*I digested BAC DNA was separated on a 1% agarose gel by pulsed-field gel electrophoresis with the following parameters, 5 to 15 s switch time at 6 V/cm, and 12.5°C in 0.5× TBE buffer for 16 h. The gel was stained with ethidium bromide, distained in water for 30 min, and photographed.

To examine the stability of half-smooth tongue sole insert DNA in the vector pECBAC1, eight selected BAC clones that had insert sizes larger than 150 kb were each inoculated into 5 ml LB medium containing 12.5 μg/ml chloramphenicol and cultured at 37°C with 250 rpm shaking overnight. Then, 1 μl of the overnight culture was inoculated in another 5 ml LB medium overnight at same growth condition. Serial cultures were made continuously for 5 days, approximately 100 generations for *E. coli*. BAC DNAs were isolated from day 1 and day 5 cultures, digested by *Hind*III, and analyzed by electrophoresis on 1% CHEF gel as above (Tao and Zhang 1998).

BAC Library Screening with Overgo Probes

To further validate the genome coverage and to assess the quality of the BAC library, 55,296 BAC clones were gridded at high density onto 22.5×22.5 cm Hybond XL+ membrane filters (Amersham, Piscataway, NJ) using the GeneTAC G3 Robotic Workstation (Genomic Solutions, Inc., USA). Each clone was double-spotted in 4×4 arrays, thereby allowing representation of 18,432 different clones per filter. Thus, the whole library was represented by three filters. The filters were placed on LB media supplemented with 12.5 mg/L of chloramphenicol and incubated overnight at 37°C. The filters were then soaked in 10% SDS for 4 min, in denaturing solution for 5 min, in neutralizing solution for 5 min, and then in 2× SSC, 0.1% SDS for 5 min, in 2× SSC for 5 min, soaked in 0.4 N NaOH for 20 min, rinsed in 5× SSC, 0.1% SDS for 20 min, in 2× SSC for 10 min, and air-dried overnight and stored at 4°C. Then, two sets of five-pooled overgo probes each were labeled with ^{32}P, hybridized to the high-density filters at 56°C and subsequently washed using the overgo hybridization protocol (Zhang 2003). The filters were then exposed to X-ray film for 1–3 days, depending on the strength of their hybridization signals. The putative positive clones were rearrayed into a small sublibrary, and filters were prepared using this sublibrary. The sublibrary filters were hybridized as above with individual probes to correlate the positive clones with the individual probes.

Acknowledgements We thank Xiaojun Zhang and Yang Zhang for technical assistance and Jieming Zhai and Jingfeng Yang for material preparation. This work was supported by grants from the State 863 High-Technology R&D Project of China (2006AA10A403) to SLC, the Shandong Genetic Improvement Key Project for Agricultural Organism to SLC, and the Taishan Scholar Project of Shandong Province to SLC.

References

Chang YL, Tao Q, Scheuring C, Meksem K, Zhang HB (2001) An integrated map of *Arabidopsis thaliana* for functional analysis of its genome sequence. Genetics 159:1231–1242

Chen SL, Li J, Deng SP, Tian YS, Wang QY, Zhuang ZM, Shan ZX, Xu JY (2007) Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol 9:273–280

Chen SL, Deng SP, Ma HY, Tian YS, Xu JY, Yang JF, Wang QY, Ji XS, Shao CW, Wang XL, Wu PF, Deng H, Zhai JM (2008) Molecular marker-assisted sex control in half-smooth tongue sole (*Cynoglossus semilaevis*). Aquaculture 283:7–12

Chen SL, Tian YS, Yang JF, Shao CW, Ji XS, Zhai JM, Liao XL, Zhuang ZM, Su PZ, Xu JY, Sha ZX, Wu PF, Wang N (2009) Artificial gynogenesis and sex determining in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol 11:243–251

Deng SP, Chen SL (2009) Molecular cloning, tissue, embryos and larvae expression analysis of Sox10 in half-smooth tongue sole, *Cynoglossus semilaevis*. Mar Genomics 1:109–114

Deng SP, Chen SL, Xu JY, Liu BW (2009) Molecular cloning, characterization and expression analysis of gonadal P450 aromatase in the half-smooth tongue-sole, *Cynoglossus semilaevis*. Aquaculture 287:211–218

Devlin RH, Nagahama Y (2002) Sex determining and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208:191–364

Devlin RH, Biagi CA, Smailus DE (2001) Genetic mapping of Y chromosomal DNA markers in Pacific salmon. Genetica 111:43–58

Dolezel J, Bartos H, Voglmayr J, Greilhuber J (2003) Nuclear DNA content and genome size of trout and human. Cytometry 51:127–128

Ezaz MT, Harvey SC, Boonphakdee C, Teale AJ, Mcandrew BJ, Penman DJ (2004) Isolation and physical mapping of sex-linked AFLP markers in Nile Tilapia (*Oreochromis niloticus L.*). Mar Biotechnol 6:435–445

Froschauer A, Korting C, Katagiri T, Aoki T, Asakawa S, Shimizu N, Schartl M, Volff JN (2002) Construction and initial analysis of bacterial artificial chromosome (BAC) contigs from the sex-determining region of the platyfish *Xiphophorus maculatus*. Gene 295:247–254

Griffiths R, Orr KJ, Adam A, Barber I (2000) DNA sex identification in the three-spined stickleback. J Fish Biol 57:1331–1334

Katagiri T, Asakawa S, Minagawa S, Shimizu N, Hirono I, Aoki T (2001) Construction and characterization of BAC libraries for three fish species: rainbow trout, carp and tilapia. Anim Genet 32:200–204

Katagiri T, Minagawa S, Hirono I, Kato K, Miyata M, Asakawa S, Shimizu N, Aoki T (2002) Construction of a BAC library for the red sea bream *Pagrus major*. Fisheries Sci 68:4942–4944

Kitano T, Takamune K, Kobayashi T, Nagahama S-I, Abe Y (1999) Suppression of P450 aromatase gene expression in sex-reversed males produced by rearing genetically female larvae at a high water temperature during a period of sex diVerentiation in the Japanese flounder (*Paralichthys olivaceus*). J Mol Endocrinol 23:167–176

Lee BY, Penman DJ, Kocher TD (2003) Identification of a sex determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. Anim Genet 34:379–383

Li Y, Hill JA, Yue GH, Chen F, Orban L (2002) Extensive search does not identify genomic sex markers in *Tetraodon nigroviridis*. J Fish Biol 61:1314–1317

Liao X, Shao CW, Tian YS, Chen SL (2007) Polymorphic dinucleotide microsatellites in tongue sole (*Cynoglossus semilaevis*). Mol Ecol Notes 7:1147–1149

Liao XL, Ma HY, Xu GB, Shao CW, Tian YS, Ji XS, Yang JF, Chen SL (2009) Construction of a genetic linkage map and mapping of a female-specific DNA marker in half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol. doi:10.1007/s10126-009-9184-3

Marchand O, Govoroun M, DCotta Carreras H, McMeel OM, Lareyre JJ, Bernot A, Laudet V, Guiguen Y (2000) DMRT1 expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. BBA-Gene Struct Expr 1493:180–187

Masakatsu W, Naoki K, Asao F, Norihiro O (2003) Construction of a BAC library for *Haplochromis chilotes*, a cichlid fish from Lake Victoria. Genes Genet Syst 78:103–105

Matsuda M (2003) Sex determining in fish: lessons from the sex-determining gene of the teleost medaka, *Oryzias latipes*. Develop Growth Differ 45:397–403

Matsuda M, Kawato N, Asakawa S, Shimizu N, Nagahama Y, Hamaguchi S, Sakaizumi M, Hori H (2001) Construction of a BAC library derived from the inbred Hd-rR strain of the teleost fish, *Oryzias latipes*. Genes Genet Syst 76:61–63

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

Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, Shimizu A, Shan Z, Haaf T, Shimizu N, Shima A, Schmid M, Schartl M (2002) A duplicated copy of *DMYT1* in the sex-determining region of the Y chromosome of the medaka, *Oryzias latipes*. Proc Natl Acad Sci USA 99:11778–11783

Peruzzi S, Chatain B, Menu B (2005) Flow cytometric determination of genome size in European seabass (*Dicentrarchus labrax*), gilthead seabream (*Sparus aurata*), thinlip mullet (*Liza ramada*), and European eel (*Anguilla anguilla*). Aquat Living Resour 18:77–81

Quiniou SM, Katagiri T, Miller NW, Wilson M, Wolters WR, Waldbieser GC (2003) Construction and characterization of a BAC library from a gynogenetic channel catfish *Ictalurus punctatus*. Genet Sel Evol 35:673–83

Ren C, Xu ZY, Sun S, Lee MK, Wu C, Zhang HB (2005) Genomic libraries for physical mapping. In: Meksem K, Kahl G (eds) The handbook of plant genome mapping: genetic and physical mapping. Wiley, Weinheim, pp 173–213

Schartl M (2004) A comparative view on sex determination in medaka. Mech Develop 121:639–645

Shao CW, Liao XL, Chen SL (2008) Isolation and characterization of microsatellite DNA loci from the southern flounder, *Paralichthys lethostigma*. Mol Ecol Resour 8:381–383

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, Smith MJ, Foster JW, Frischauf AM, Lovell-Badge R, Goodfellow PN (1990) A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346:240–244

Takayuki K, Shuichi A, Ikuo H, Takashi A, Nobuyoshi S (2000) Genomic bacterial artificial chromosome library of the Japanese flounder *Paralichthys olivaceus*. Mar Biotechnol 2:571–576

Tao Q, Zhang HB (1998) Cloning and stable maintenance of DNA fragments over 300 kb in *Escherichia coli* with conventional plasmid-based vectors. Nucleic Acids Res 26:4901–4909

Tao Q, Chang YL, Wang J, Chen H, Islam-Faridi MN, Scheuring C, Wang B, Stelly DM, Zhang HB (2001) Bacterial artificial chromosome-based physical map of the rice genome constructed by restriction fingerprint analysis. Genetics 158:1711–1724

Thorsen J, Zhu BL, Frengen E, Osoegawa K, Jong PJ, Koop BF, Davidson WS, Høyheim B (2005) A highly redundant BAC library of Atlantic salmon (*Salmo salar*): an important tool for salmon projects. BMC Genomics 6:50–58

Tiersch TR, Chandler RW, Kallman K, Wachtel SS (1989) Estimation of nuclear DNA content by flow cytometry in fishes of the genus *Xiphophorus*. Comp Biochem Physiol 94:465–468

Volff JN, Kondo M, Schartl M (2003) Medaka dmY/dmrt1Y is not the universal primary sex-determining gene in fish. Trends Genet 19:196–199

Volff JN, Nanda I, Schmid M, Schartl M (2007) Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sex Dev 1:85–99

Wang S, Xu P, Thorsen J, Zhu B, de Jong PJ, Waldbieser G, Kucuktas H, Liu Z (2007) Characterization of a BAC library from channel catfish *Ictalurus punctatus*: Indications of high levels of chromosomal reshuffling among teleost genomes. Mar Biotechnol 9:701–11

Wang X, Zhang Q, Sun X, Chen Y, Zhai T, Zhuang W, Qi J, Wang Z (2009) Fosmid library construction and initial analysis of end sequences in female half-smooth tongue sole (*Cynoglossus semilaevis*). Mar Biotechnol 11:236–242

Whitaker HA, McAndrew BJ, Taggart JB (2006) Construction and characterization of a BAC library for the European sea bass *Dicentrarchus labrax*. Anim Genet 37:526

Xu P, WangS LL, Peatman E, Somridhivej B, Thimmapuram J, Gong G, Liu Z (2006) Channel catfish BAC-end sequences for marker development and assessment of syntenic conservation with other fish species. Anim Genet 37:321–326

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

Yokoi H, Kobayashi T, Tanaka M, Nagahama Y, Wakamatsu Y, Takeda H, Araki K, Morohashi K, Ozato K (2002) Sox9 in a teleost fish, medaka (*Oryzias latipes*): evidence for diversified function of Sox9 in gonad differentiation. Mol Reprod Dev 63:5–16

Zhang HB (2003) BAC protocol. In: Laboratory for plant genomics and GENEfinder genomics resource. http://hbz7.tamu.edu

Zhang HB, Wu C (2001) BAC as tools for genome sequencing. Plant Physiol Biochem 39:1–15

Zhang HB, Zhao X, Ding X, Paterson AH, Wing RA (1995) Preparation of megabase-size DNA from plant nuclei. Plant J 7:175–184

Zhang HB, Woo S, Wing RA (1996) BAC, YAC and Cosmid library construction. In: Foster G, Twell D (eds) Plant gene isolation: principles and practice. Wiley, England, pp 75–99

Zhang Y, Zhang X, Scheuring CF, Zhang HB, Huan P, Li F, Xiang J (2008) Construction and characterization of two bacterial artificial chromosome libraries of Zhikong scallop, *Chlamys farreri* Jones et Preston, and identification of BAC clones containing the genes involved in its innate immune system. Mar Biotechnol 10:358–65

Zhang X, Zhang Y, Scheuring C, Zhang HB, Huan P, Wang B, Liu C, Li F, Liu B, Xiang J (2009) Construction and characterization of a bacterial artificial chromosome (BAC) library of Pacific white shrimp, *Litopenaeus vannamei*. Mar Biotechnol. doi:10.1007/s10126-009-9209-y

Zhou LQ, Yang AG, Liu XZ, Du W, Zhuang ZM (2005) The karyotype analysis of *Cynoglossus semilaevis* in China. J Fish China (in Chinese with English abstract) 29:417–419

Characterization of the *cyp19a1a* gene from a BAC sequence in half-smooth tongue sole and analysis of its conservation among teleosts

Abstract

The *cyp19a1a* gene encoding the aromatase plays a key role in sex differentiation of gonad. We report here the first BAC sequence of half-smooth tongue sole (*Cynoglossus semilaevis*) containing intact *cyp19a1a* gene, and characterization of the common conservation of *cyp19a1a* gene as well as comparison of the gene synteny among teleosts. The BAC is 107,367 bp in size with an overall GC content of 43.44% and transposable elements of 4.38%. A total of 9 genes including 7 functional genes and 2 hypothetical genes were predicted. The *cyp19a1a* gene of all tested teleosts has 9 exons and 8 introns, and some conserved potential binding sites flanking the transcriptional start site. The expression pattern was also similar during the ovarian differentiation. Besides, gene syntenic analysis reveals a conserved gene cluster *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA* among teleosts. These findings suggest that *cyp19a1a* among teleosts are not only similar in the genomic structure but also conserved in function, even though the genomic environment of *cyp19a1a* in tongue sole is not universal in teleosts, reflecting the special evolution of tongue sole *cyp19a1a* after the split of the other teleosts.

Keywords

BAC; Cyp19a1a; Conservation; Gene synteny; Half-smooth tongue sole

Introduction

Cytochrome P450 aromatase (P450arom) is a terminal enzyme of the aromatase complex that catalyzes the synthesis of androgen to estrogen, which are generally regarded as exogenous hormone responsible for reproductive physiology in vertebrate, especially sex determination and differentiation in teleosts. For instance, inhibiting aromatase activity can result in various degrees of masculinization including complete phenotypic and functional males in some teleost fishes, such as Chinook salmon *Oncorhynchus tshawytscha* (Piferrer et al. 1994), Nile tilapia *Oreochromis niloticus*, rainbow trout *Oncorhynchus mykiss* (Guiguen et al. 1999), Japanese flounder *Paralichthys olivaceus* (Kitano et al. 2000), and zebrafish *Danio rerio* (Uchida et al. 2004). Thus, P450 aromatase is considered to be a key steroid and aromatase for ovarian differentiation and sex reversal occurred at the critical developmental stage of some teleost species.

The *cyp19* gene encoding P450arom is believed to occur as a single copy in the genome of most mammals and the tissue-specific expression was achieved by alternative splicing and/or different promoter usage in these species (Simpson et al. 1993). In human, for example, a single *cyp19* gene locates on the chromosome 15, containing 9 coding exons and 1 noncoding exon. The multiple tissue-specific first exons in human *cyp19* give birth to different transcripts in various tissues and responsible for the tissue specific expression (Mendelson et al. 1987; Simpson et al. 1993). By contrast, most teleost species have two separate and distinctive genes, designated *cyp19a1a* and *cyp19a1b*, resulting from the fish-specific genome

duplication (FSGD) (Callard and Tchoudakova 1997). The *cyp19a1a* gene encodes the P450aromA was predominantly present in gonad while the other is mostly brain specific gene encoding P450aromB. Besides, for most teleost species, the initiation codon of *cyp19a1a* was located in the exon 1 (Chang et al. 2005; Galay-Burgos et al. 2006; Kanda et al. 2006; Zhang et al. 2008), whereas the start codon was resided in the exon 2 for *cyp19a1b*, which is consistent with the organization of the mammalian *cyp19* gene (Chang et al. 2005; Dalla Valle et al. 2005). Both of *cyp19* genes have different tissue specificities, expression pattern, enzymatic activities and regulation modes, as reported in zebrafish, *Danio rerio* (Kishida and Callard 2001; Trant et al. 2001); European sea bass, *Dicentrarchus labrax* (Blazquez and Piferrer 2004); Nile tilapia, *Oreochromis niloticus* (Chang et al. 2005); Atlantic halibut, *Hippoglossus hippoglossus* (Matsuoka et al. 2006; van Nes et al. 2005); common carp, *Cyprinus carpio* (Barney et al. 2008); red-spotted grouper, *Epinephelus akaara* (Huang et al. 2009); yellowtail clownfish, *Amphiprion clarkii* (Kobayashi et al. 2010) and catfish, *Clarias gariepinus* (Sridevi et al. 2012). A detailed example of existence of two *cyp19* genes was first reported in goldfish. *Cyp19a1b* was discovered as a 3 kb transcript in brain with a high abundance by hybridation while a 1.9 kb cDNA encoding the *cyp19a1a* was isolated through a stepwise PCR cloning strategy in goldfish. These two genes have only 62% identity, which is evidently less than identity of *cyp19a1a* between goldfish and other teleosts (68%-72%), but presumptive functional domains are highly conserved (Tchoudakova and Callard 1998).

In half-smooth tongue sole (*Cynoglossus semilaevis*), as in other teleost species, the

cyp19 gene also has two isoforms encoding brain and gonadal aromatases, respectively (Deng et al. 2009a). The tongue sole has a distinct trait that the female grow two to four times faster than males; and more important the natural population exhibits the skewed sex ratios with almost 70% offspring being male due to the spontaneously sex reversal (Ji et al. 2010). It is undesirable for aquaculture that the high ratio of male can result in overall reduction of output and thus leads to heavy economic loss. Although the tongue sole employs a ZW/ZZ system of genetic sex determination (GSD), the phenotypic sex was influenced by both genetic and environmental factors including temperature, sex hormone and other environmental factors. Treatment with high temperature (28℃) during a critical developmental stage on offspring leads to about 70% sex-reversed ZW male individuals (Deng et al. 2007; Tian et al. 2011). The same phenomenon was also observed by the effects of 80 μg/L methyl-testosterone (MT) which was widely considered as inhibitor of aromatizable estrogens, suggesting that the aromatase was involved in sex differentiation of tongue sole (Chen et al. 2008).

The importance of aromatase in sex differentiation, particularly in teleost fishes, makes it impressive to understand the common features of *cyp19* gene among independent teleost lineages, such as conserved regulation mechanism and genomic environment. To date, however, most of documents about the *cyp19* gene were focused on the expression pattern, promoter activity and their function in sexual differentiation for a species. Only a few studies have addressed structurally and functionally conservation of *cyp19* gene among teleosts (Castro et al. 2005; Guiguen

et al. 2010).

In present study, we reported the sequencing and characterization of BAC clone containing a full-length *cyp19a1a* gene of tongue sole and, consequently comparation of *cyp19a1a* gene including the genomic structure, expression pattern and gene synteny, suggesting a relatively conserved pattern of *cyp19a1a* gene among teleosts.

Material and Methods

Tongue sole samples

All the developmental stages of half-smooth tongue sole were collected from Shandong Huanghai aquaculture Co., Ltd. For the early stages including the 4 days post hathching (dph), 8 dph, 25 dph and 48 dph of tongue sole fries, the undifferentiated gonads situating at the anterior part of the whole intestine were removed and then cut into several parts using a fine forceps under a stereoscopic microscope. We then assessed the suitability of isolated gonad parts using the sex specific marker and gonad specific gene. Gonads (five to ten fries and three adults) of same age group and same sex were pooled and stored at −80°C for RNA isolation.

BAC sequence and assembly

A BAC clone Hind025N15 containing the intact *cyp19a1a* gene was previously identified by Overgo hybridization, originating from BAC libraries of tongue sole (Shao et al. 2010). BAC DNA purified by Qiagen Q-Tip100 was sheared by nebulization to an average size of 0.6 kb. After end-filling, DNA fragments were size-fractionated and cloned into the plasmid of pUC18. The resulting plasmids were

electroporated into DH5α competent cells (TAKARA). Clones were picked into 96-well trays filled with LB culture medium, grown for 12 h and frozen until needed. Approximately 1,992 clones were sequenced with the ABI PRISM Dye Terminator Cycle Sequencing Ready Reaction kit (Applied Biosystems) on ABI 3730 sequencing analyzer. Sequences were manually edited, aligned, and assembled using Sequencer software (PHRED and PHRAP) (Ewing et al. 1998) (http://www.phrap.org/).

Analysis of sequence data

Repetitive elements were identified by RepeatMasker (version 3.2.6) against the Repbase TE library (version 2008-08-01). Then the BAC sequence was masked by RepeatMasker to perform gene predictions using two *de novo* prediction software programs: Genscan (Salamov and Solovyev 2000) and Augustus (Stanke and Waack 2003), trained by human genome annotation. The predicted genes were checked by the domain search software interpro (http://www.ebi.ac.uk/interpro/). A gene with significant homology [E less than (e-20)] to a known protein is classified according to the protein name as human while the other gene without significant homology to any protein is classified as hypothetical protein. The homology searches were carried out using the Basic Blast program, at http://www.ncbi.nlm.nih.gov/blast/.

The synteny comparisons across different teleost species *Takifugu rubripes*, *Tetraodon nigroviridis*, *Oryzias latipes*, *Gasterosteus aculeatus* and *Danio rerio* were carried out *in silico* by means of reciprocal BLAST 'best-hit' searches using the ENSEMBL database (http://www.ensembl.org). Orthologous gene location data were

retrieved from the genomes of these teleosts, and were arranged by reference to the order of BAC genes.

For *cyp19a1a* gene, the deduced amino acid sequences of tongue sole, together with *cyp19a1a* amino acid sequences reported for other teleost species were aligned by Clustal W version 1.6 using default settings (Thompson et al. 1994). Furthermore, the conserved domain was identified by GENEDOC program version 2.6.02 (http://www.psc.edu/biomed/genedoc/). A phylogenetic tree was constructed with the neighbor joining method using Mega 5 (Tamura et al. 2011) and the *cyp19* gene from chicken and human served as an out group to root the tree. The exon-intron boundaries as well as the intron sequences were determined by comparing the genomic sequence with the deduced amino acid sequence using Genewise2-2-0 (Birney et al. 2004). The transcription start site and putative transcriptional factor binding sites were identified by MatInspector (http://www.genomatix.de).

Expression of cyp19a1a during the sex differentiation stage

Total RNA of gonads at different developmental stages was isolated and reversely transcribed as described previously (Dong et al. 2011; Chen et al. 2001). A pair of primers (F: GGTGAGGATGTGACCCAGTGT; R: ACGGGCTGAAATCGCAAG) for qRT-PCR analysis was designed using the Primer3 program. The final PCR reactions contained 0.4 mM of each primer; 10 μL SYBR Green (Invitrogen) and 80 ng template of cDNA. qRT-PCR was performed on ABI PRISM 7500 Real-Time PCR System using Hotstart *Taq* polymerase (Qiagen) in a final volume of 20 μL and *β*-actin gene was used as internal reference. All reactions were subjected to: 95 for 35 s followed by 40 cycles at 95℃ for 5 s, 60℃ 34 s and then 95℃ 15 s , 60℃ for

1 min and 95℃ 15 s. Melting curve analysis was applied to all reactions to ensure homogeneity of the reaction product. The result was analyzed using 7500 System SDS Software.

Results

Characterization of BAC sequence

The assembly of the whole BAC Hind025N15 sequence (GenBank accession no. JQ003881) is 107,367 bp from approximately 1,589 useful reads with average length of 548 bp, covering about 7.69 times length of whole BAC. Two gaps in region unable to be fully sequenced and/or assembled would be result from richness of repeat sequence. The GC content in the whole BAC clone is 43.44% and the sequence consists of totally about 4.38% of transposable elements (TE). DNA transposons are predominant and make up 2.71% of the entire BAC clone covering 6,390 bp, whereas retrotransposons account for 2.42% of the BAC length including 1.21% LINE (2,850 bp)，0.13 % SINE (307 bp) and 2,553 bp length of LTR alone accounts for 1.08% of all transposable elements (Table 1).

Totally nine genes are predicted from the BAC Hind025N15 sequence using two programs, Genscan and Augustus. Seven of them have homologous genes by BLASTP, BLASTN and BLASTX to conserved regions while the other two genes (*G0001* and *G0002*) were defined as hypothetical genes in the order of *CP19A*, *IDH3A*, *PKH4B*, *SL9A5*, *FHOD3*, *G0001*, *CEBPG*, *CEBPA* and *G0002*. The average gene size of the 9 genes in this BAC is 5,469 bp and the average gene density is 1 gene/11,929 bp. The first one is the target gene, *cyp19a1a*, which is 2,457 bp resided

from 3,854 to 6,311 on Hind025N15 (Table 1). The next three functional genes are length of 2,448 bp (*IDH3A*), 13,532 bp (*PKH4B*) and 6,150 bp (*SL9A5*) long, respectively (Table 1). The *IDH3A*, isocitrate dehydrogenase 3 (NAD+) alpha, is an enzyme that catalyzes the oxidative decarboxylation of isocitrate to alpha-ketoglutarate (alpha-KG), requiring nicotinamide adenine dinucleotide (NAD) as a cofactor (Huh et al. 1996). The *PKH4B* is the short name of pleckstrin homology domain-containing family G member 4B involving the process of rho protein signal transduction (Nagase et al., 2001). The *SL9A5,* solute carrier family 9 member 5, takes part in the pH regulation to eliminate acids generated by active metabolism (Klanke et al. 1995). The fifth gene *FHOD3,* formin homology 2 domain containing 3, is the largest gene with length of 17,716 bp, and might play a role in the regulation of the actin cytoskeleton (Iskratsch et al. 2010). The others two small genes (*CEBPG* and *CEBPA*) encode CCAAT/enhancer binding protein (C/EBP) which is a DNA-binding protein that recognizes two different motifs, the CCAAT homology common to many promoters and the enhanced core homology common to many enhancers (Hattori et al. 2003).

Genomic organization of *cyp19a1a*

The exon/intron organization of the tongue sole *cyp19a1a* gene was determined by the BAC sequence compared with cDNA of *cyp19a1a* which was identified by homology-based cloning method. Also, the genomic structure of other teleosts *cyp19a1a* was confirmed by the whole genome sequence and cDNA of *cyp19a1a*. It

is of interest that comparison of the exon/intron structure among teleost species revealed a common conservation with 9 exons and 8 introns, which were inserted at exactly the same positions as those found in human (Tanaka et al. 1995) (Figure 1b). All donor and acceptor sites of these introns were GT and AG, respectively, following the GT/AG rule. The total length of the introns of *cyp19a1a* in different teleosts was distributed through 702 bp to 13,968 bp while the total length of exons was resided from 1,402 to 1,578 bp, suggesting that the gene size of *cyp19a1a* was determined by the length of introns (Figure 1b). The examples are *Takifugu* and *Tetraodon cyp19a1a* gene span only 2,224 and 2,256 bp, being much smaller than the zebrafish *cyp19a1a* gene (15,519 bp), as a result of extremely small introns with a total of 758 bp and 702 bp, respectively.

The deduced amino acid sequence of the tongue sole *cyp19a1a* was aligned with other teleost *cyp19a1a* homologues (Figure 2) and a phylogenetic tree was further constructed using human and chicken *cyp19a1a* as out group. The resulting tree demonstrates that the two paralogs are distributed two main clades corresponding to brain and ovary aromatase present in teleosts (Figure 3). Sequence alignment demonstrated five conserved putative function domains, namely the membrane spanning region, substrate binding region, distal helix I region, steroid binding region and heme-binding helix region (Figure 2). Comparison of the deduced amino acid sequence revealed that the tongue sole *cyp19a1a* gene shared a high degree of amino acid sequence identity with medaka (77.7%), *Takifugu* (76.4%), *Tetraodon* (75.1%), stickleback (76.3%), Japanese flounder (77.0%), southern flounder (76.1%) and barfin

flounder (75.0%), respectively. But it had a relatively lower level of identity with zebrafish (68.1%) and evidently higher than degrees of overall sequence identities with human (58.4%) and chicken (60.2%), respectively. Besides, only 64.7% identity was indicated when full-length *cyp19a1a* and *cyp19a1b* in tongue sole were compared.

To further explore the conservation of the teleost *cyp19a1a*, we compared 1000 bp upstream of the translation start sites, in detail, in zebrafish, tongue sole, medaka, *Takifugu*, *Tetraodon* and stickleback. We identified a highly conserved TATA box and several potential regulatory elements including C/EBP (CCAAT binding factors), FKHD (Fork head domain factors), SF1F (Vertebrate steroidogenic factor), SORY (SOX/SRY-sex/testis determinig and related HMG box factors), CREB (cAMP-responsive element binding proteins), EREF (Estrogen response elements) and GATA (GATA binding factors) among six different teleost species (Figure 1c). However, the transcriptional binding site of *SF-1* which is an important transcription factor to regulate *cyp19* genes in human ovary was not identified in the 5' flanking region of tongue sole and *Tetraodon cyp19a1a*. Also the EREF which is the binding site of estrogen receptor was not detected on the *Tetraodon cyp19a1a*. Besides, there are other potential binding sites which were not conserved in most teleost species were detected by MatInspector such as an arylhydrocarbon receptor (AhR) recognition site, a nerve growth factor inducible-B protein and a doublesex and mab-3 related transcription factor 1(DMRT).

Expression of *cyp19a1a* gene of tongue sole

Expression of *cyp19a1a* during developmental stages from 4 dph to 2 years old in tongue sole was quantified by real time PCR. The result showed that the expression of *cyp19a1a* gene was increased significantly in gonad samples of developing juveniles for the first time at 70 dpf with the drastically dimorphic expression pattern between male and female. Then the *cyp19a1a* expression have increased evidently up to 160 dpf which is the stage of highest expression level both of male and female. Expression decrease thereafter in females and persist at high level in mature ovary, while the expression was sustained at lower level in male relatively to the female (Figure 4).

Gene synteny

Seven protein-coding genes of BAC Hind025N15 in tongue sole were further mapped onto other teleost genomes by BLASTP and the location of orthologous in respective genome were retrieved. The results showed that the gene order and orientation in the *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA* cluster are conserved among five teleosts (medaka, *Tetraodon*, *Takifugu*, stickleback and tongue sole) except that the *SL95A* and *CEBPA* were missing in medaka and stickleback, respectively (Figure 1a). Although the zebrafish have the same genes with other teleost species in this region, it had not been exhibited obvious conservation because the intensively interchromosomal arrangement caused gene disorder including the pair of *FHOD3* and *PKH4B* and the pair of *CEBPA* and *CEBPG*. No significant similarity to *IDH3A* was identified in the *Tetraodon*, *Takifugu*, stickleback and zebrafish genome. But in

the medaka and tongue sole genome, it was detected in the corresponding region where the synteny has not been well kept. The gene *cyp19a1a* (*CP19A*) was distributed on the same chromosome with conserved cluster in *Tetraodon* (Chromosome 5) and stickleback (Group II), but it kept at least 4 Mb distance to the proximate gene *PKH4B* of conserved cluster. In contrast, it was detected on the different chromosomes and/or scaffolds with conserved cluster in *Takifugu* (Scaffold 336 versus Scaffold 1), medaka (Ultracontig 49 versus Chromosome 3) and zebrafish (Chromosome 7 versus Chromosome 18) (Figure 1a).

Discussion

BAC clones are highly useful for identifying specific gene organization across relatively long chromosomal distances and thus facilitating the comparative analysis among different species, especially for the non-model species with no whole genome sequence available (Yasukochi et al. 2011). We reported here the first BAC clone containing the entire *cyp19a1a* gene that exhibited the special function during the embryonic development of tongue sole and further developed the comparison analysis among different lineage of teleost species, revealed the common conservation of this gene.

Cyp19a1a, encoded P450aromatase catalyzing the synthesis of estrogens, is a well studied gene found in species ranging from teleost to mammalian (Diotel et al. 2010). In most teleosts, the structure and function of the *cyp19a1a* gene has been well conserved during evolution (Callard and Tchoudakova 1997; Castro et al. 2005). The

genomic structure of the tongue sole *cyp19a1a* gene containing 9 exons and 8 introns was almost the same as those of other teleost species. It is noteworthy that although the gene size variated among different teleost species, there are minor variations in the sizes of the CDS region (1,402 bp-1,578 bp) while large dissimilarities are witnessed in the introns of different species, especially a paradigmatic example of comparison between zebrafish (13,968 bp) and *Tetraodon* (702 bp). As expected, the sizes of the introns differ across species, and some are extremely long, mainly accounts for much of the difference in genome size among teleost species.

A cross-species comparison of *cyp19a1a* sequences among teleosts indicates five regions of greatest conservation including membrane spanning region (Ⅰ), substrate binding region (Ⅱ), distal helix Ⅰ region (Ⅲ), steroid binding region (Ⅳ) and heme-binding helix region (Ⅴ). Conceivably, the steroid binding region, denoted aromatase-specific domain, is functionally essential in all tested species, while Ⅰ, Ⅱ, Ⅲ, and Ⅴ have evolved conserved function within the teleosts (Huang et al. 2009). Interestingly, similar to gold fish *cyp19a1a* (Callard and Tchoudakova 1997), the tongue sole *cyp19a1a* gene, despite having only about 64.7% identity with its ortholog *cyp19a1b*, is possess high identity with other teleost species (68.1%-77.7%), reflecting their conserved expression pattern in ovary and brain, respectively.

DNA sequences denoted *cis*-acting elements that regulate expression of the *cyp19a1a* gene by binding transcription factors are located within a region spanning -1000 bp that flank the transcription start site. The teleost species with the genome sequence available, has allowed us to examine the manner and conservation in which the

cyp19a1a promoter to control the level of transcription initiation, which largely determines gene expression in different teleost species. The major findings are that the TATA box is consistently conserved with respect to position, and that the most highly conserved regulatory elements of EREF, C/EBP, FKHD, SF1F, SORY, CREB and GATA sequences clearly indicate a common mechanism of regulation in tested teleost species. Imperfect consensus estrogen response element (ERE) were identified in *Tetraodon* sequence, as in the case of the gonadal aromatase promoter from the orange-spotted grouper sequence (Zhang et al. 2012). This would be interesting to investigate whether there is another regulatory model to control the gonadal development in *Tetraodon* or just a mistake for prediction by unconsensus of software. As a binding site of transcriptional factor SF1 that mediated *cyp19a1a* expression during vitellogenesis in Nile tilapia (Yoshiura et al. 2003), SF1F was also not detected in the promoter of tongue sole and *Tetraodon cyp19a1a*. Further investigations are required to clarify this issue by experiment and it would be the same reason with ERE in *Tetraodon*. FOXL2 is known as the earliest sexually dimorphic marker in ovarian development and it is essential for granulosa cell differentiation and ovarian maintenance (Baron et al. 2004). Several other studies implicated that FOXL2 is a candidate regulator of the gonadal aromatase gene as shown in rainbow trout (Kanda et al. 2006) and Nile tilapia (Wang et al. 2007). The other conserved potential sites including C/EBP, SORY, CREB and GATA have previously been described in gonadal aromatase promoters from tilapia, European seabass, medaka (Chang et al. 2005; Dalla Valle et al. 2002; Tanaka et al. 1995). Together, these studies contribute

to our understanding of *cyp19a1a* transcriptional regulation mechanism, but how transcriptional modulation and chromatin structure remodelling are controlled by the promoter is still poorly understood and as yet there is no complete analysis of the *cyp19a1a* promoter.

In general, the expression level of *cyp19a1a* is consistent with the level of estrogen which is essential for ovarian development. Previously study revealed that the tongue sole *cyp19a1a* transcript was mainly expressed in the ovary of female (ZW♀) and a low expression level in gonad of pseudo-male (ZW♂) produced by both MT immersion and high temperature treatment (Deng et al. 2009a). This result indicates that *cyp19a1a* is indispensable for ovarian differentiation in tongue sole. Thus, in the present study, we determined the expression pattern of *cyp19a1a* during the stages of sex determination and differentiation in tongue sole. The sexual dimorphisim between female and male at 70 dph which the stage of gonadal differentiation accompanying the appearance of primary oocytes indicated again that the pivotal role of *cyp19a1a* in ovarian differentiation in tongue sole. The specific over-expression of *cyp19a1a* in vitellogenic follicles during oogenesis has been also observed in many fish species including Nile tilapia (Sudhakumari et al. 2005), rainbow trout (Guiguen et al. 1999; Vizziano et al. 2007), European seabass (Blazquez et al. 2008) and medaka (Nakamoto et al. 2006; Patil and Gunasekera 2008). The similar expression pattern of *cyp19a1a* gene among teleost species suggests their involvement in ovarian differentiation.

Comparative genomic analysis has been used extensively to identify similar genomic

features and to establish gene co-linearity to some extent among different species with close taxonomic relationships (Deng et al. 2009b). Our analysis of gene synteny between tongue sole BAC genes and their counterparts in other five teleost species presents a conserved cluster *PKH4B-SL9A5-FHOD3-CEBPG-CEBPA*, suggesting a microcolinearity between teleost genomes. However, we detected a distinct alternation in manner of *cyp19a1a* gene which was not in the conserved syntenic regions where it distribute far away from common conserved gene cluster in *Tetraodon* and stickleback genome or even locate on different chromosomes and/or scaffolds in medaka, *Takifugu* and zebrafish genome. Since the scaffolds containing the counterparts of tongue sole genes in medaka and *Takifugu* had not mapped onto the chromosome, we lack sufficient information to conclude whether the *cyp19a1a* gene and the conserved cluster distribute linearly on the same chromosome. In order to illustrate this issue, we analyzed the upstream and downstream of *cyp19a1a* gene in other five teleost species based on the available genome sequences. Unexpectedly, two conserved gene clusters distributing around *cyp19a1a* gene were detected. One is the *DMXL2-GLDN-CP19A-TNFAIP8L3* cluster which was detected in syntenic region of medaka, stickleback, *Takifugu*, *Tetraodon* and zebrafish. Another cluster is *SCG3-LYSMD2-TMOD2-LEO1-MAPK6-USP50*，which was conserved in medaka, *Takifugu*, *Tetraodon* and zebrafish but not on the syntenic region of stickleback (data not shown). Furthermore, we traced back to human and found that all the genes of two conserved clusters exhibit linear arrangement on human chromosome 15, indicating that *cyp19a1a* gene have a common genomic environment (data not shown). Based on

data that exists, the possibilities to explain the different manner of *cyp19a1a* gene arrangement occurred in the syntenic region of tongue sole may be caused by the insertions and/or translocations and/or duplication of *cyp19a1a* gene after the split of the other teleost species. Taken together, despite the discrepancies exists in the syntenic region between tongue sole and other teleost species, all the teleost species still shared highly conserved gene content.

In conclusion, the BAC clone Hind025N15 we report is the first large genomic sequence of tongue sole. Analysis of the TEs, GC content and gene organization of BAC sequence give a profile of the whole genome sequence of tongue sole at a close firstly. Comparison analysis revealed that the *cyp19a1a* gene had the common conserved features such as the exon/intron pattern, the potential binding sites upstream of the transcriptional start site, and its function in the ovarian differentiation among teleosts. Although the *cyp19a1a* of tongue sole was not followed the common synteny, the detection of gene cluster (*PKH4B-SL9A5-FHOD3-CEBPG-CEBPA*) are conserved in all tested fish species. These characteristics will make the BAC clone an excellent starting platform for future functional studies of *cyp19a1a* gene. Additional whole genome sequencing efforts will provide further insight into the evolutionary history of the *cyp19a1a* in tongue sole.

Acknowledgements:

This work was supported by grants from Special Fund for Agro-scientific Research in the Public Interest (200903046) of China, National Nature Science Foundation of

China (31130057, 31072202 and 41006107), and Taishan Scholar Project Fund of Shandong of China.

References

Barney ML, Patil JG, Gunasekera RM, et al. 2008. Distinct cytochrome P450 aromatase isoforms in the common carp (*Cyprinus carpio*): sexual dimorphism and onset of ontogenic expression. Gen Comp Endocrinol, 156: 499-508

Baron D, Cocquet J, Xia X, et al. 2004. An evolutionary and functional analysis of FoxL2 in rainbow trout gonad differentiation. J Mol Endocrinol, 33：705-715

Birney E, Clamp M, Durbin R. 2004. GeneWise and Genomewise. Genome Res, 14: 988-995

Blazquez M, Gonzalez A, Papadaki M, et al. 2008. Sex-related changes in estrogen receptors and aromatase gene expression and enzymatic activity during early development and sex differentiation in the European sea bass (*Dicentrarchus labrax*). Gen Comp Endocrinol, 158: 95-101

Blazquez M, Piferrer F. 2004. Cloning, sequence analysis, tissue distribution, and sex-specific expression of the neural form of P450 aromatase in juvenile sea bass (*Dicentrarchus labrax*). Mol Cell Endocrinol, 219: 83-94

Callard GV, Tchoudakova A. 1997. Evolutionary and functional significance of two CYP19 genes differentially expressed in brain and ovary of goldfish. J Steroid Biochem Mol Biol, 61: 387-392

Castro LF, Santos MM, Reis-Henriques MA. 2005. The genomic environment around

the Aromatase gene: evolutionary insights. BMC Evol Biol, 5: 43

Chang X, Kobayashi T, Senthilkumaran B, et al. 2005. Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). Gen Comp Endocrinol, 141: 101-115

Chen SL, Hong YH, Scherer S J, et al. 2001. Lack of ultraviolet-light inducibility of the medaka fish (Oryzias latipes) tumor suppressor gene. Gene, 264: 197–203

Chen SL, Deng SP, Ma HY, et al. 2008. Molecular marker-assisted sex control in half-smooth tongue sole (*Cynoglossus semilaevis*). Aquaculture, 283: 7-12

Dalla Valle L, Lunardi L, Colombo L, et al. 2002. European sea bass (*Dicentrarchus labrax* L.) cytochrome P450arom: cDNA cloning, expression and genomic organization. J Steroid Biochem Mol Biol, 80: 25-34

Dalla Valle L, Toffolo V, Vianello S, et al. 2005. Genomic organization of the CYP19b genes in the rainbow trout (*Oncorhynchus mykiss* Walbaum). J Steroid Biochem Mol Biol, 94: 49-55

Deng SP, Chen SL, Xu JY, et al. 2009a. Molecular cloning, characterization and expression analysis of gonadal P450 aromatase in the half-smooth tongue-sole, *Cynoglossus semilaevis*. Aquaculture, 287: 211-218

Deng SP, Chen SL, Tian YS, et al. 2007. Gonadal differentiation and effects of temperature on sex determination in half-smooth tongue sole, *Cynoglossus semilaevis*. Journal of fishery science of china (in Chinese), 14: 714-719

Deng ZY, Li Y, Xia R, et al. 2009b. Structural analysis of 83-kb genomic DNA from *Thellungiella halophila*: sequence features and microcolinearity between salt cress

and Arabidopsis thaliana. Genomics, 94: 324-332

Diotel N, Le Page Y, Mouriec K, et al. 2010. Aromatase in the brain of teleost fish: expression, regulation and putative functions. Front Neuroendocrinol, 31: 172-192

Dong XL, Chen SL, Ji XS, et al. 2011. Molecular cloning, characterization and expression analysis of Sox9a and Foxl2 genes in half-smooth tongue sole (*Cynoglossus semilaevis*). Acta Oceanologica Sinica, 30: 68-77

Ewing B, Hillier L, Wendl MC, et al. 1998. Base-calling of automated sequencer traces using phred. I. Accuracy assessment. Genome Res, 8: 175-185

Galay-Burgos M, Gealy C, Navarro-Martin L, et al. 2006. Cloning of the promoter from the gonadal aromatase gene of the European sea bass and identification of single nucleotide polymorphisms. Comp Biochem Physiol A Mol Integr Physiol, 145: 47-53

Guiguen Y, Baroiller JF, Ricordel MJ, et al. 1999. Involvement of estrogens in the process of sex differentiation in two fish species: the rainbow trout (*Oncorhynchus mykiss*) and a tilapia (*Oreochromis niloticus*). Mol Reprod Dev, 54: 154-162

Guiguen Y, Fostier A, Piferrer F, et al. 2010. Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. Gen Comp Endocrinol, 165: 352-366

Hattori T, Ohoka N, Inoue Y, et al. 2003. C/EBP family transcription factors are degraded by the proteasome but stabilized by forming dimer. Oncogene, 22: 1273-1280

Huang W, Zhou L, Li Z, et al. 2009. Expression pattern, cellular localization and promoter activity analysis of ovarian aromatase (Cyp19a1a) in protogynous

hermaphrodite red-spotted grouper. Mol Cell Endocrinol, 307: 224-236

Huh TL, Kim YO, Oh IU, et al. 1996. Assignment of the human mitochondrial NAD+ -specific isocitrate dehydrogenase alpha subunit (IDH3A) gene to 15q25.1-->q25.2 by in situ hybridization. Genomics, 32: 295-296

Iskratsch T, Lange S, Dwyer J, et al. 2010. Formin follows function: a muscle-specific isoform of FHOD3 is regulated by CK2 phosphorylation and promotes myofibril maintenance. J Cell Biol, 191: 1159-1172

Ji XS, Chen SL, Ma HY, et al. 2010. Natural sex reversal of female *Cynoglossus semilaevis* in rearing populations. Journal of fisheries of China (in Chinese), 34: 322-335

Kanda H, Okubo T, Omori N, et al. 2006. Transcriptional regulation of the rainbow trout CYP19a gene by FTZ-F1 homologue. J Steroid Biochem Mol Biol, 99: 85-92

Kishida M, Callard GV. 2001. Distinct cytochrome P450 aromatase isoforms in zebrafish (*Danio rerio*) brain and ovary are differentially programmed and estrogen regulated during early development. Endocrinology, 142: 740-750

Kitano T, Takamune K, Nagahama Y, et al. 2000. Aromatase inhibitor and 17alpha-methyltestosterone cause sex-reversal from genetical females to phenotypic males and suppression of P450 aromatase gene expression in Japanese flounder (*Paralichthys olivaceus*). Mol Reprod Dev, 56: 1-5

Klanke CA, Su YR, Callen DF, et al. 1995. Molecular cloning and physical and genetic mapping of a novel human Na+/H+ exchanger (NHE5/SLC9A5) to chromosome 16q22.1. Genomics, 25: 615-622

Kobayashi Y, Horiguchi R, Miura S, et al. 2010. Sex- and tissue-specific expression of P450 aromatase (cyp19a1a) in the yellowtail clownfish, *Amphiprion clarkii*. Comp Biochem Physiol A Mol Integr Physiol, 155: 237-244

Matsuoka MP, van Nes S, Andersen O, et al. 2006. Real-time PCR analysis of ovary- and brain-type aromatase gene expression during Atlantic halibut (*Hippoglossus hippoglossus*) development. Comp Biochem Physiol B Biochem Mol Biol, 144: 128-135

Mendelson CR, Evans CT, et al. 1987. Regulation of aromatase in estrogen-producing cells. J Steroid Biochem, 27: 753-757

Nagase T, Kikuno R, Ohara O. 2001. Prediction of the coding sequences of unidentified human genes. XXI. The complete sequences of 60 new cDNA clones from brain which code for large proteins. DNA Res, 8: 179-187

Nakamoto M, Matsuda M, Wang DS, et al. 2006. Molecular cloning and analysis of gonadal expression of Foxl2 in the medaka, *Oryzias latipes*. Biochem Biophys Res Commun, 344: 353-361

Patil JG, Gunasekera RM. 2008. Tissue and sexually dimorphic expression of ovarian and brain aromatase mRNA in the Japanese medaka (*Oryzias latipes*): implications for their preferential roles in ovarian and neural differentiation and development. Gen Comp Endocrinol, 158: 131-137

Piferrer F, Zanuy S, Carrillo M, et al. 1994. Brief treatment with an aromatase inhibitor during sex differentiation causes chromosomally female salmon to develop as normal, functional males. J Exp Zool, 270: 255-262

Salamov AA, Solovyev VV. 2000. Ab initio gene finding in Drosophila genomic DNA. Genome Res, 10: 516-522

Shao CW, Chen SL, Scheuring CF, et al. 2010. Construction of two BAC libraries from half-smooth tongue sole *Cynoglossus semilaevis* and identification of clones containing candidate sex-determination genes. Mar Biotechnol, 12: 558-568

Simpson ER, Mahendroo MS, Means GD, et al. 1993. Tissue-specific promoters regulate aromatase cytochrome P450 expression. J Steroid Biochem Mol Biol, 44: 321-330

Sridevi P, Chaitanya RK, Dutta-Gupta A, et al. 2012. FTZ-F1 and FOXL2 up-regulate catfish brain aromatase gene transcription by specific binding to the promoter motifs. Biochim Biophys Acta, 1819: 57-66

Stanke M, Waack S. 2003. Gene prediction with a hidden Markov model and a new intron submodel. Bioinformatics, 19 Suppl 2: ii215-225

Sudhakumari CC, Senthilkumaran B, Kobayashi T, et al. 2005. Ontogenic expression patterns of several nuclear receptors and cytochrome P450 aromatases in brain and gonads of the Nile tilapia *Oreochromis niloticus* suggests their involvement in sex differentiation. Fish Physiol Biochem, 31: 129-135

Tamura K, Peterson D, Peterson N, et al. 2011. MEGA5: molecular evolutionary genetics analysis using maximum likelihood, evolutionary distance, and maximum parsimony methods. Mol Biol Evol, 28: 2731-2739

Tanaka M, Fukada S, Matsuyama M, et al. 1995. Structure and promoter analysis of the cytochrome P-450 aromatase gene of the teleost fish, medaka (*Oryzias latipes*). J

Biochem, 117: 719-725

Tchoudakova A, Callard GV. 1998. Identification of multiple CYP19 genes encoding different cytochrome P450 aromatase isozymes in brain and ovary. Endocrinology, 139: 2179-2189

Thompson JD, Higgins DG, Gibson TJ. 1994. CLUSTAL W: improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. Nucleic Acids Res, 22: 4673-4680

Tian YS, Wang D, Xu Y, et al. 2011. Effects of rearing temperature on growth and sex determination in the half-smooth tongue sole (*Cynoglossus semilaevis*). Journal of fisheries of china (in Chinese), 35: 176-182

Trant JM, Gavasso S, Ackers J, et al. 2001. Developmental expression of cytochrome P450 aromatase genes (CYP19a and CYP19b) in zebrafish fry (*Danio rerio*). J Exp Zool, 290: 475-483

Uchida D, Yamashita M, Kitano T, et al. 2004. An aromatase inhibitor or high water temperature induce oocyte apoptosis and depletion of P450 aromatase activity in the gonads of genetic female zebrafish during sex-reversal. Comp Biochem Physiol A Mol Integr Physiol, 137: 11-20

van Nes S, Moe M, Andersen O. 2005. Molecular characterization and expression of two cyp19 (P450 aromatase) genes in embryos, larvae, and adults of Atlantic halibut (*Hippoglossus hippoglossus*). Mol Reprod Dev, 72: 437-449

Vizziano D, Randuineau G, Baron D, et al. 2007. Characterization of early molecular

sex differentiation in rainbow trout, *Oncorhynchus mykiss*. Dev Dyn, 236: 2198-2206

Wang DS, Kobayashi T, Zhou LY, et al. 2007. Foxl2 up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. Mol Endocrinol, 21: 712-725

Yasukochi Y, Tanaka-Okuyama M, Kamimura M, et al. 2011. Isolation of BAC clones containing conserved genes from libraries of three distantly related moths: a useful resource for comparative genomics of Lepidoptera. J Biomed Biotechnol, 2011: 165894

Yoshiura Y, Senthilkumaran B, Watanabe M, et al. 2003. Synergistic expression of Ad4BP/SF-1 and cytochrome P-450 aromatase (ovarian type) in the ovary of Nile tilapia, *Oreochromis niloticus*, during vitellogenesis suggests transcriptional interaction. Biol Reprod, 68: 1545-1553

Zhang WM, Lu HJ, Jiang HY, et al. 2012. Isolation and characterization of cyp19a1a and cyp19a1b promoters in the protogynous hermaphrodite orange-spotted grouper (*Epinephelus coioides*). Gen Comp Endocrinol, 175: 473-487

Zhang Y, Zhang WM, Yang HY, et al. 2008. Two cytochrome P450 aromatase genes in the hermaphrodite ricefield eel *Monopterus albus*: mRNA expression during ovarian development and sex change. J Endocrinol, 199: 317-331

Table 1 Characterization of BAC sequence in half-smooth tongue sole

BAC	**size**	**coverage**	**Total reads No.**	**Useful reads no.**	**Average length**	**Contig no.**	**GC (%)**
Hind025N15	107367	7.69	1992	1589	548	3	43.44
Repeat type	**DNA**	**LINE**	**SINE**	**LTR**	**Other**	**Unknown**	**Total**
Length (bp)	6390	2850	307	2553	0	0	10289
% genome	2.72	1.21	0.13	1.08	0	0	4.38
Protein gene	**CP19A**	**IDH3A**	**PKH4B**	**SL95A**	**FHOD3**	**CEBPG**	**CEBPA**
Gene length (bp)	2457	2448	13532	6150	17716	491	935

Figure Legends

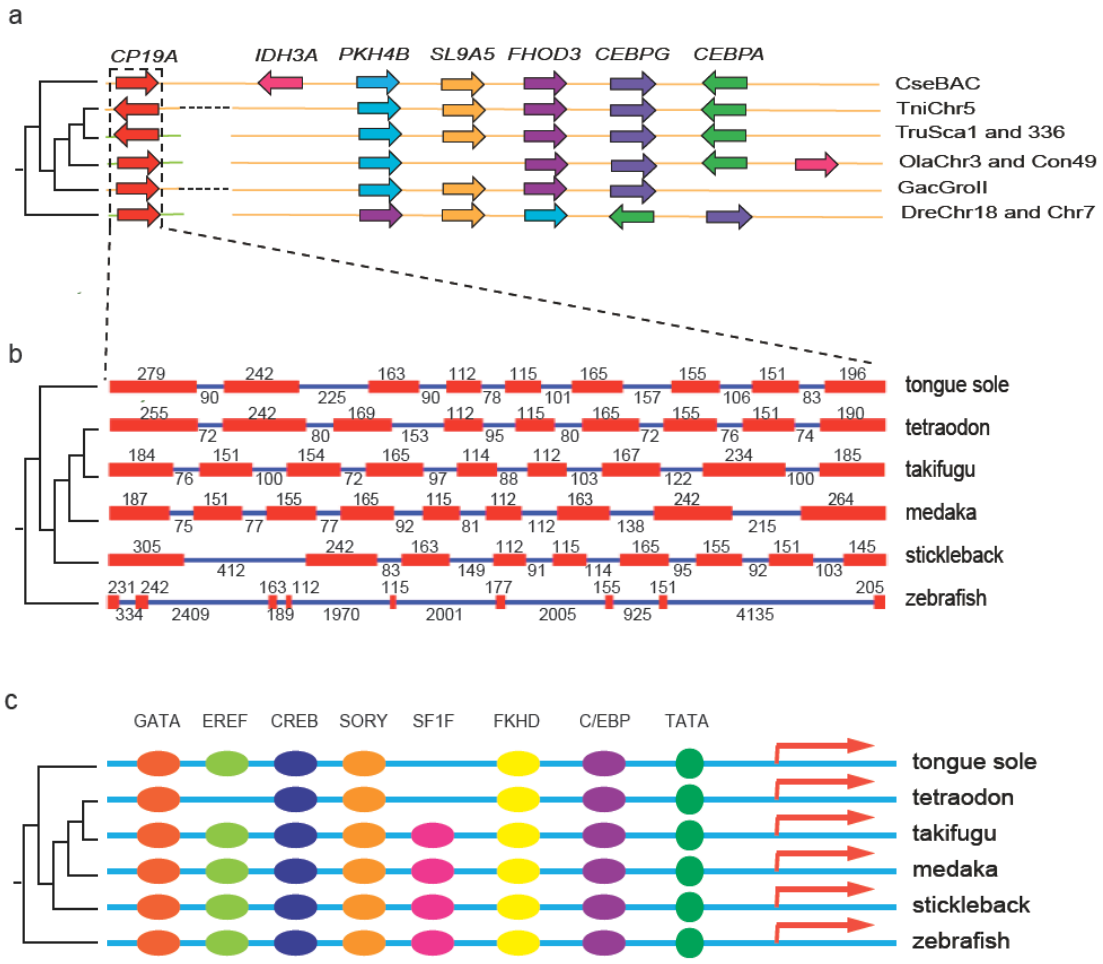

Fig. 1 Comparison of *cyp19a1a* gene among six different teleost species. **a, Teleost gene synteny**. Genomic organization and gene synteny comparisons across teleost for genes annotated from tongue sole BAC clone Hind025N15. Genes are depicted by colored arrows and arrowhead indicates orientation for each gene. Gene names are indicated above the tongue sole genes, and orthologs across species are depicted in the same colors. The black dashed lines from the genomic region of *Takifugu* and stickleback indicate that the *cyp19a1a* and the other orthologs had a big distance. For the genomic region of *Tetraodon*, zebrafish and medaka, the orthologs distributed on the different chromosomes or scaffolds indicating by two color lines. CseBAC: BAC sequence of tongue sole; TniChr5: Chromosome 5 of *Takifugu*; TruScaf1 and 336: scaffold 1 and scaffold 336 of *Tetraodon*; OlaChr3 and Con49：Chromosome 3 and ultracontig 49 of medaka; GacGroII: Group II of stickleback;

DreChr18 and Chr7: Chromosome 18 and chromosome 7 of zebrafish. **b, Exon and intron structure of *cyp19a1a***. Gene exon and intron structures are depicted for six teleost species. Exons for *cyp19a1a* are represented by red boxes connected by introns depicted as blue lines. Exon and intron lengths are labelled in base pairs; exon lengths are indicated directly below each exon, and intron lengths are indicated directly above each intron. **c, Transcription factor binding sites in the *cyp19a1a* promoter region**. The green circle represent the conserved TATA box and the other colored circles with the name above the tongue sole BAC sequence represent the potential binding sites predicted by MatInspector. The red arrows indicate the transcriptional start sites. Transcription factor binding sites are absolutely conserved in different teleost species but the order and spacing base pairs are poorly conserved and thus the schematic map represent the conservation of promoter region was depicted after artificial contraposition.

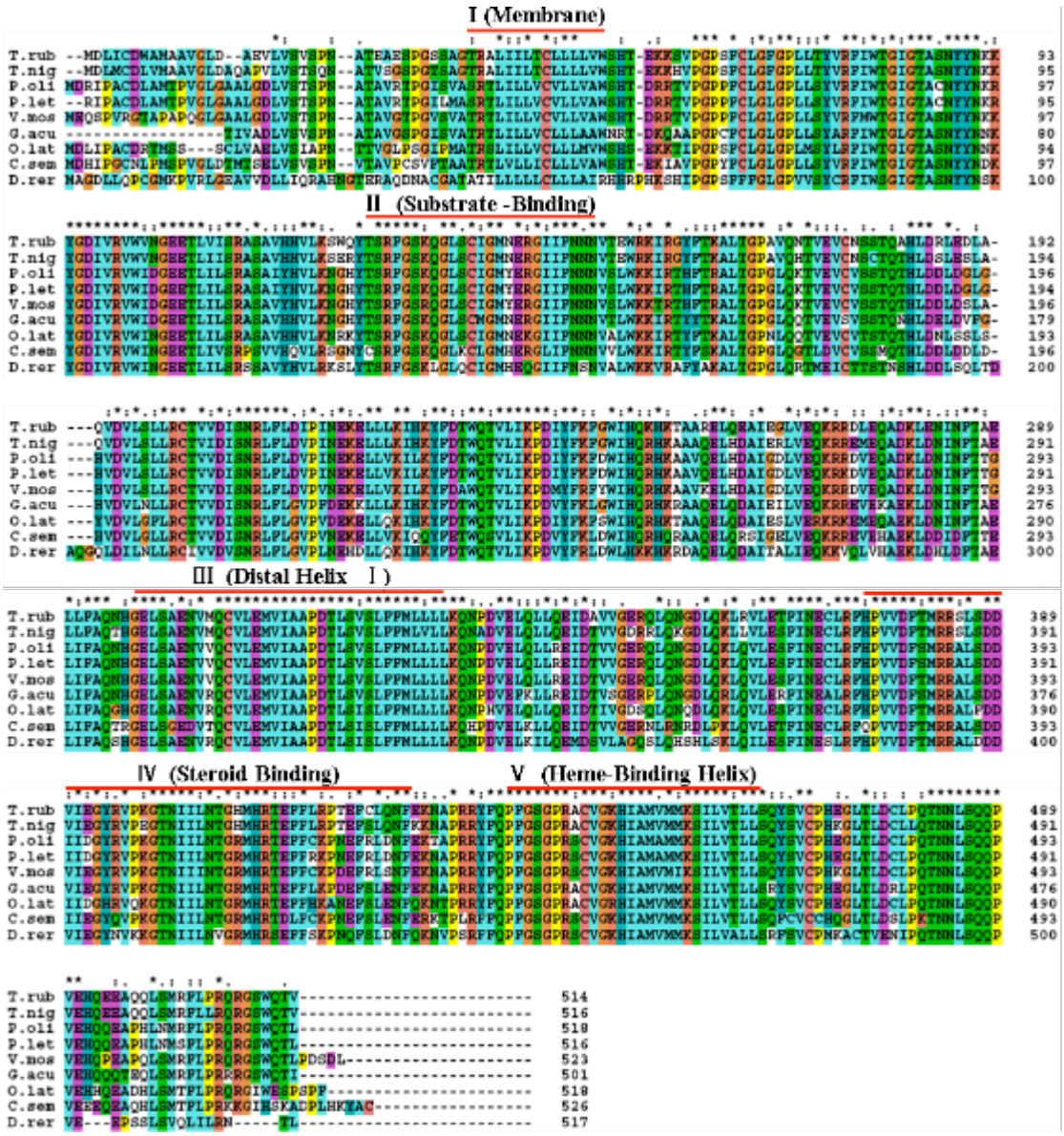

Fig. 2 Amino acid sequence alignment of *cyp19a1a* proteins from different teleost species *T.rub*, *Takifugu rubripes* (BAF93506); *T.nig, Tetraodon nigroviridis* (CAF99837)*;P.oli, Paralichthys olivaceus* (BAA74777.1); *P.let, Paralichthys lethostigma* (AAX55671.1); *V.mos, Verasper moseri* (ACI04549); *G.acu, Gasterosteus aculeatus* (ACN70015); *O.lat*, *Oryzias latipes* (BAA11656); *C.sem*, *Cynoglossus semilaevis* (ABL74474) and *D.rer*, *Danio rerio* (AAK00643). The five conserved regions are coverd by red line and marked with Roman numerals I–V. Identity is indicated by star, and gaps used to maximize the alignment are shown by dashes.

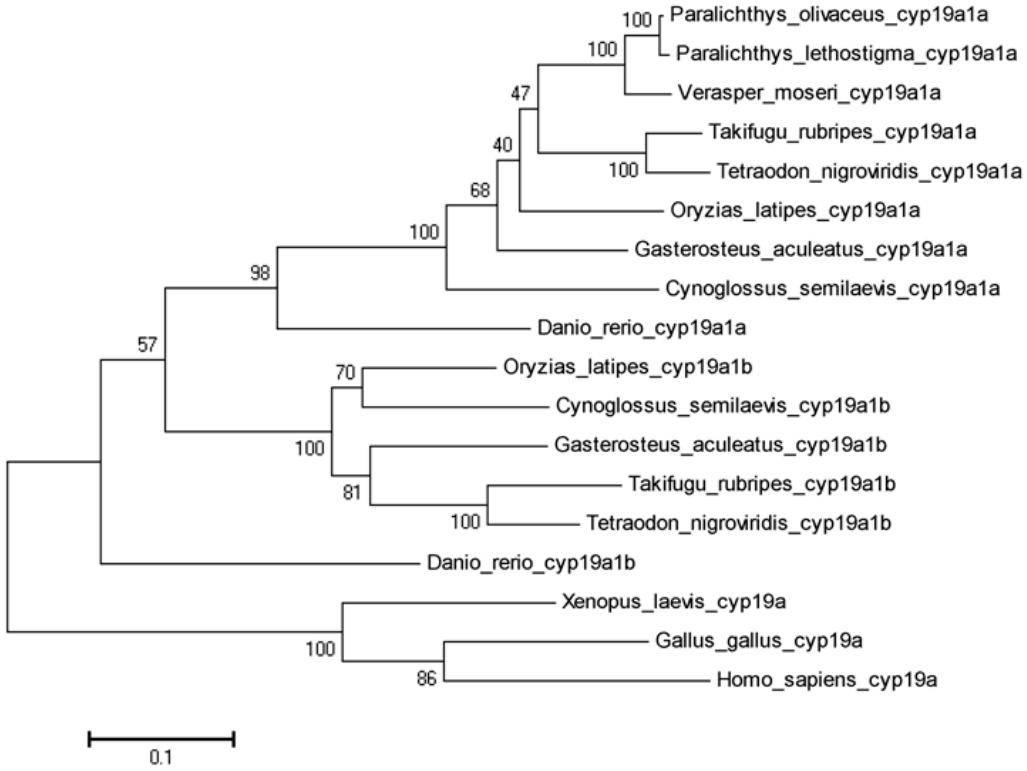

Fig. 3 Phylogenetic tree of vertebrate aromatase genes. The scientific names of the organism followed by the gene name are shown. Distances are used to construct the phylogenetic tree and bootstrap values based on 1000 resampling replicates. The jade-green region represents the cyp19a1a gene clade. The accession number of different organisms were followed by *Takifugu rubripes*_cyp19a1a (BAF93506), *Takifugu rubripes*_cyp19a1b (BAF93507), *Tetraodon nigroviridis*_cyp19a1a (CAF99837), *Tetraodon nigroviridis*_cyp19a1b (CAG05537), *Oryzias latipes*_cyp19a1a (BAA11656), *Oryzias latipes*_cyp19a1b (AAP83449), *Gasterosteus aculeatus*_cyp19a1a (ACN70015), *Gasterosteus aculeatus*_cyp19a1b (ENSGACP00000007910), *Cynoglossus semilaevis*_cyp19a1a (ABL74474), *Cynoglossus semilaevis*_cyp19a1b (ABM90641), *Paralichthys olivaceus* (BAA74777.1), *Paralichthys lethostigma* (AAX55671.1), *Verasper moseri* (ACI04549), *Danio rerio*_cyp19a1a (AAK00643), *Danio rerio*_cyp19a1b (AAK00642), *Homo sapiens* (AAR37047) and *Gallus gallus* (AAA48738).

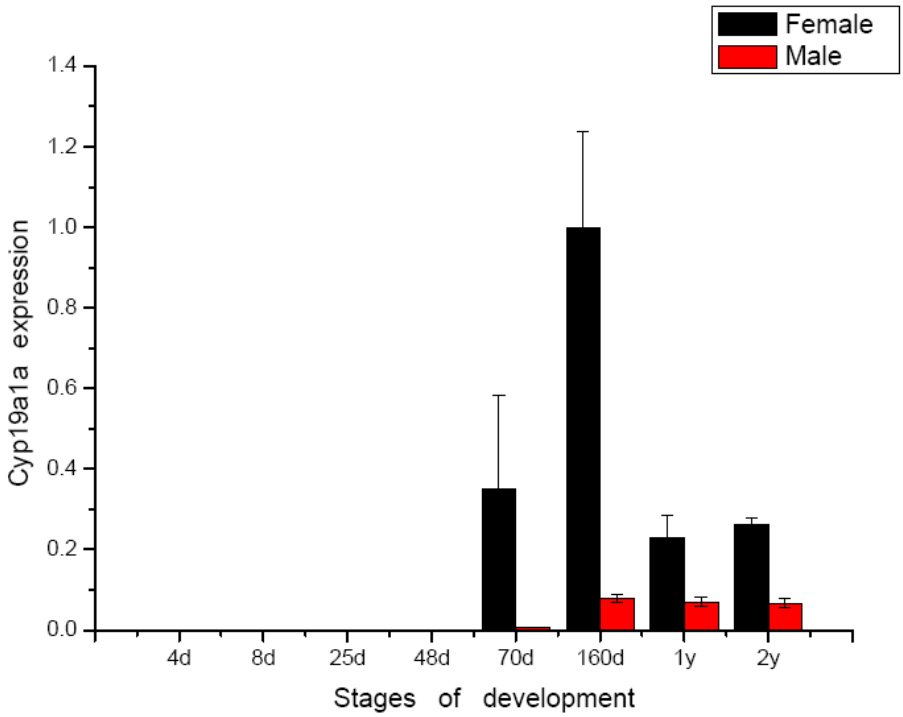

Fig. 4 Reverse transcription polymerase chain reaction (RT-PCR) analysis of *cyp19a1a* during developmental stages in female (red bar) and male (black bar) of tongue sole. Vertical bars mean ± standard error (SE) (n = 3).

Chapter Ⅲ Epigenetic inheritance of sex reversal in half-smooth tongue sole

Shao CW, Li QY, Chen SL, Zhang P, Lian JM, Hu QM, Sun B, Jin LJ, Liu SS, Wang ZJ, Zhao HM, Jin ZH, Liang Z, Li YZ, Zheng QM, Zhang Y, Wang J, Zhang GJ. Epigenetic modification and inheritance in sexual reversal of fish. Genome Research, 2014, 24: 604-615.

Epigenetic modification and inheritance in sexual reversal of fish

Changwei Shao,[1,8] Qiye Li,[2,3,4,5,8] Songlin Chen,[1,8,9] Pei Zhang,[2] Jinmin Lian,[2] Qiaomu Hu,[1] Bing Sun,[1] Lijun Jin,[2] Shanshan Liu,[1] Zongji Wang,[2,3] Hongmei Zhao,[2] Zonghui Jin,[2] Zhuo Liang,[1] Yangzhen Li,[1] Qiumei Zheng,[2] Yong Zhang,[2] Jun Wang,[2,6,7] and Guojie Zhang[2,4,9]

[1]Yellow Sea Fisheries Research Institute, CAFS, Key Lab for Sustainable Development of Marine Fisheries, Ministry of Agriculture, Qingdao 266071, China; [2]China National Genebank, BGI-Shenzhen, Shenzhen 518083, China; [3]School of Bioscience and Bioengineering, South China University of Technology, Guangzhou 510006, China; [4]Centre for Social Evolution, Department of Biology, University of Copenhagen, DK-2100 Copenhagen, Denmark; [5]Centre for GeoGenetics, Natural History Museum of Denmark, University of Copenhagen, 1350 Copenhagen K, Denmark; [6]Department of Biology, University of Copenhagen, DK-2200 Copenhagen, Denmark; [7]King Abdulaziz University, Jeddah 22254, Saudi Arabia

Environmental sex determination (ESD) occurs in divergent, phylogenetically unrelated taxa, and in some species, co-occurs with genetic sex determination (GSD) mechanisms. Although epigenetic regulation in response to environmental effects has long been proposed to be associated with ESD, a systemic analysis on epigenetic regulation of ESD is still lacking. Using half-smooth tongue sole (*Cynoglossus semilaevis*) as a model—a marine fish that has both ZW chromosomal GSD and temperature-dependent ESD—we investigated the role of DNA methylation in transition from GSD to ESD. Comparative analysis of the gonadal DNA methylomes of pseudomale, female, and normal male fish revealed that genes in the sex determination pathways are the major targets of substantial methylation modification during sexual reversal. Methylation modification in pseudomales is globally inherited in their ZW offspring, which can naturally develop into pseudomales without temperature incubation. Transcriptome analysis revealed that dosage compensation occurs in a restricted, methylated cytosine enriched Z chromosomal region in pseudomale testes, achieving equal expression level in normal male testes. In contrast, female-specific W chromosomal genes are suppressed in pseudomales by methylation regulation. We conclude that epigenetic regulation plays multiple crucial roles in sexual reversal of tongue sole fish. We also offer the first clues on the mechanisms behind gene dosage balancing in an organism that undergoes sexual reversal. Finally, we suggest a causal link between the bias sex chromosome assortment in the offspring of a pseudomale family and the transgenerational epigenetic inheritance of sexual reversal in tongue sole fish.

[Supplemental material is available for this article.]

Mechanisms of sex determination can be broadly divided into two major categories: genetic sex determination (GSD) and environmental sex determination (ESD). In organisms with GSD (often based on sex chromosomes), the primary sex of an individual is determined at the moment of fertilization by heritable genetic elements that differ between the sexes, whereas organisms with ESD do not possess a primary sex at fertilization and have their sex induced during ontogeny (Bull 1983; Stelkens and Wedekind 2010; Matsumoto and Crews 2012). Generally, the sex of GSD species will be fixed for life, but in some GSD species, the primary sex can be altered during development, such that individuals can develop into the opposite sex without changing their genotype. This phenomenon is known as environmental sex reversal (ESR) (Stelkens and Wedekind 2010) and is relatively common in insects (Vance 1996; Narita et al. 2007), fishes (Devlin and Nagahama 2002), amphibians (Wallace et al. 1999), and reptiles (Quinn et al. 2007). Diverse environmental triggers for ESR have been documented, including abiotic (e.g., temperature, pH, hormones) and biotic (e.g., crowding, pathogens) factors, of which temperature is the most broadly studied (Kato et al. 2011). Currently, artificial induction of sex reversal is widely used in aquaculture by exposure of fish to exogenous steroids or different rearing temperatures in order to produce the preferred sex (Stelkens and Wedekind 2010). ESD lacks the flexibility of ESR, but may be adaptive because it can preferentially produce the sex that benefits most from the best possible hatching environment in terms of future reproductive success (Janzen 1995; Warner and Shine 2005, 2008). However, a common concern is that ESD systems may become a burden under rapid environmental change, such as global warming and ocean acidification, because it might result in nonadaptive sex ratios that are skewed toward predominantly male or female offspring (Ospina-Alvarez and Piferrer 2008; Mitchell and Janzen 2010; Stelkens and Wedekind 2010). Understanding the mechanisms by which genotype and environment interact to control early sex determination is therefore both of fundamental interest

[8]These authors contributed equally to this work.
[9]Corresponding authors
E-mail zhanggj@genomics.cn
E-mail chensl@ysfri.ac.cn
Article published online before print. Article, supplemental material, and publication date are at http://www.genome.org/cgi/doi/10.1101/gr.162172.113.

© 2014 Shao et al. This article is distributed exclusively by Cold Spring Harbor Laboratory Press for the first six months after the full-issue publication date (see http://genome.cshlp.org/site/misc/terms.xhtml). After six months, it is available under a Creative Commons License (Attribution-NonCommercial 3.0 Unported), as described at http://creativecommons.org/licenses/by-nc/3.0/.

24:604–615 Published by Cold Spring Harbor Laboratory Press; ISSN 1088-9051/14; www.genome.org

and of relevance for domesticated livestock and natural population conservation.

Half-smooth tongue sole, *Cynoglossus semilaevis*, is a commercially valuable flatfish that is widely distributed in Chinese coastal waters. In contrast to many other teleost species, classical karyotype analysis and artificial gynogenesis tests have revealed that this species employs a female heterogametic sex determination system (ZW♀/ZZ♂) and has clear sexual dimorphism, with females growing much faster and reaching final body sizes that are 2–4 times those of males (Zhuang et al. 2006; Chen et al. 2009). However, despite the primary determination of sex by chromosome inheritance, ~14% of ZW genetic females are known to be sex-reversed to phenotypic males when reared under normal conditions (22°C) (Chen et al. 2014). Furthermore, we have recently demonstrated that exposure to relatively high temperatures (28°C) during a sensitive developmental period early in life can increase the sex reversal rate of ZW genetic females to ~73% (Chen et al. 2014), suggesting that the sexual fate of tongue sole can be overridden by environmental factors. Interestingly, these so-called sex reversed "pseudomales" are fertile and can mate with normal females to produce viable offspring. Furthermore, the F1 generation continues to exhibit an extremely high sex reversal rate (~94%), even when reared in normal conditions (22°C) (Chen et al. 2014). This implies that the ability of sex reversal can be inherited. Therefore, with its complex sex determination system governed by the interaction between genetic determination and environmental factors, the tongue sole is an excellent model with which to understand the molecular mechanism of sex determination and ESR in fishes. We recently sequenced the genome of this fish and generated a 477-Mb high-quality assembly with up to 93.3% (445 Mb) sequences anchored to 22 chromosomes and 21,516 annotated protein-coding genes (Chen et al. 2014). Comparative genomic analysis revealed that the sex chromosome pair of the tongue sole was derived from the same ancestral vertebrate proto-chromosome that gave rise to the W and Z chromosomes of birds. We also proposed that the sex of tongue sole is determined through a Z-encoded mechanism, with male development being driven by the Z-encoded gene *dmrt1* (*doublesex and mab-3 related transcription factor 1*), which is also the male determining gene in birds (Chen et al. 2014). Thus the coexistence of GSD and ESD (in fact juvenile ESR) within this species, coupled with the availability of its full genome sequence, provides an opportunity to comprehensively investigate the gene expression and gene regulation mechanisms mediating sex determination and ESR at the whole genome level.

It has been proposed that the epigenetic regulation of gene expression during gonad differentiation could provide a basis for adaptive sex reversal in some GSD organisms (Manolakou et al. 2006; Matsumoto and Crews 2012), something that has been demonstrated to date in a few fish examples. For example, experiments in sea bass, *Dicentrarchus labrax*, suggested that the high temperature incubation during the early development stage increased the DNA methylation level of the promoter region of *cyp19a1a* (*cytochrome P450, family 19, subfamily A, polypeptide 1a*), leading to suppressed *cyp19a1a* expression and male development (Navarro-Martín et al. 2011). However, previous studies have been limited to individual genes, and a comprehensive assessment of epigenetic regulation of sex reversal is still lacking. Here, we performed thorough analyses of whole genome DNA methylome and transcriptome data on half-smooth tongue sole, including gonadal samples collected from different sexes (normal female, normal male, pseudomale) of parental fish and F1 generations of pseudomale, and we offer several novel insights on the roles of epigenetic regulation in sexual reversal of tongue sole fish.

Results

Gonadal DNA methylomes of half-smooth tongue sole

To assess the gonadal DNA methylome patterns across different sexual types of tongue sole, we carried out BS-seq (Xiang et al. 2010) on bisulfite-converted DNA extracted from adult gonads of parental females (P-ZWf), parental pseudomales (P-ZWm), F1 pseudomales (F1-ZWm), and females (F1-ZWf) from a cross between a parental pseudomale and a normal female. We also sampled "normal" male individuals (ZZm) as a control for the ZZm DNA methylation pattern. In addition, we obtained RNA-seq data (Chen et al. 2014) for corresponding samples to quantify gene expression (Fig. 1C; Supplemental Table S1).

In total, ~171 Gb methylome data were produced, which yielded an average depth of 22× per strand for each sample (Supplemental Table S2) with ~90% of genomic cytosines (Cs) being covered by at least two unique reads (Supplemental Table S3). We identified an average of 14.5 million methylated cytosines (mCs) in each sample (see Methods), accounting for 9% of Cs and 86% of CpGs in the reference genome (Supplemental Table S4). More than 99% of mCs were in the CpG context (Fig. 1D; Supplemental Fig. S3A); therefore we decided to focus solely on CpG sites for subsequent analyses. Analysis of the methylation status of CpGs in various genomic elements showed that the majority of CpGs in exons, introns, and repeats were hypermethylated (methylation level >0.75), whereas hypomethylated CpGs (methylated level <0.25) were relatively enriched in gene promoters and CpG islands (CGIs) (Fig. 1E; Supplemental Fig. S3B). Furthermore, ovary harbored more partially methylated CpGs (methylation level between 0.25 and 0.75) compared with testis (Fig. 1E; Supplemental Fig. S3B). As expected, gene promoter methylation was negatively correlated with gene expression (Fig. 1F). In general, the characteristics of DNA methylation as well as the genomic features related to DNA methylation (e.g., CpG o/e, CGIs, and *dnmt* genes) in tongue sole are similar to those observed in zebrafish (Supplemental Figs. S1, S2, S3; Jiang et al. 2013; Potok et al. 2013).

Genomic methylation patterns are consistent with gonadal differentiation

To obtain an overview of DNA methylation patterns for the five types of gonads, we first examined chromosome-wide methylation levels. We found that the methylation patterns of pseudomale testes were similar to those of normal male testes, and all three testis samples were clearly distinguished from the ovary samples by hierarchical clustering analysis (Fig. 2A). Moreover, the overall methylation levels were consistently enhanced by ~10% in testes compared to ovaries, except for the W chromosome (Fig. 2A; Supplemental Table S5). The relatively high methylation levels of the W chromosome in both ovaries and testes were probably associated with its high repeat content (Supplemental Fig. S4).

We subsequently investigated the regions that displayed significant methylation changes among samples by performing pairwise comparisons to identify differentially methylated regions (DMRs) (see Methods). In total, only 60 kb of DMRs (0.015% of the genome) were identified between P-ZWf and F1-ZWf, indicating their high concordance in DNA methylation patterns. Similarly, only 160 kb of DMRs (0.040% of the genome) were identified be-

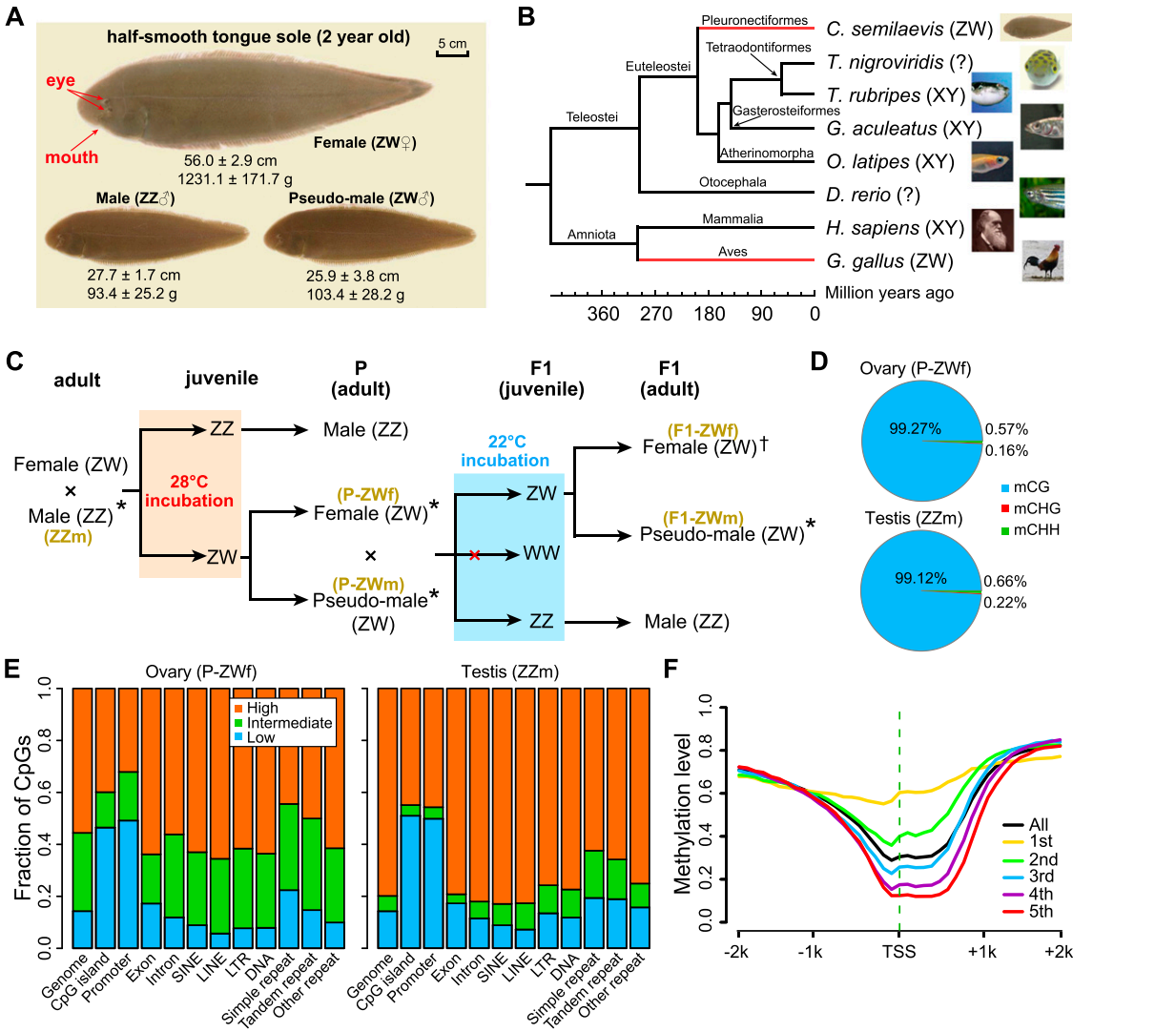

Figure 1. Morphology, phylogeny, and DNA methylation of half-smooth tongue sole. (*A*) Photos of a female, a normal male, and a pseudomale at 2 yr of age. (*B*) Genome-based phylogenetic positions and sex chromosome systems of half-smooth tongue sole and other vertebrates. Red lines show that the sex chromosome pairs of tongue sole and chicken were derived from the same ancestral vertebrate proto-chromosome pairs in spite of their distant evolutionary relationship (Chen et al. 2014). (?) No sex chromosomes have been identified so far. (*C*) Experimental design: The offspring from a normal male (ZZ) and a female (ZW) were exposed to 28°C during the sensitive developmental period, which induced the development of genetic females (ZW) into pseudomales. One of these pseudomales was subsequently crossed with one normal female to produce F1 pseudomales and females. F1 offspring carrying WW sex chromosomes do not exist, as sperms with W instead of Z are not viable. (*) Samples used for both BS-seq and RNA-seq; (†) samples only used for BS-seq. Brown letters in parentheses indicate the symbols for corresponding samples used throughout this paper. (*D*) Percentage of mCs in the CG, CHG, and CHH contexts. (*E*) Fraction of CpG in low (methylation level less than 0.25), intermediate (between 0.25 and 0.75), and high (greater than 0.75) methylation levels in different genomic elements. (*F*) Methylation profile along transcript start sites (TSS) of genes in different expression quintiles. The first quintile is the lowest and the fifth is the highest. Dashed green line indicates the location of TSS. *B* and *C* are modified from Figures S16 and 3a in Chen et al. (2014), respectively.

tween P-ZWm and F1-ZWm, a figure close to that observed between ZWm and ZZm (~225 kb; 0.056% of the genome), suggesting that methylation patterns of the three testis samples were also highly similar. In contrast, we identified an average of ~15 Mb of DMRs (4% of the genome) between testes and ovaries, orders of magnitude more than that between two ovary samples or among three testis samples (Fig. 2B). Taken together, our data suggest that the genome-wide methylation patterns of genetic females have been accurately shaped to the patterns of normal males after phenotypic sex reversal.

Interestingly, up to 86% of the DMRs between P-ZWm and ovaries were maintained in F1-ZWm (F1 pseudomale of P-ZWm) (Fig. 2C; Supplemental Fig. S5), indicating that the overall change in methylation status of the genome after sex reversal had been inherited or re-established by the next generation. Furthermore, ~95% of these maintained DMRs were also observed between ZZm and ovaries (Fig. 2C; Supplemental Fig. S5), implying that the DNA methylation changes after sex reversal were associated with gonad differentiation in tongue sole.

Gonadal DMGs are associated with development, morphogenesis, and reproduction

Because the overall methylation patterns of pseudomale testes (P-ZWm and F1-ZWm) were very similar to those of normal male testes, we next focused on DMRs where methylation levels were

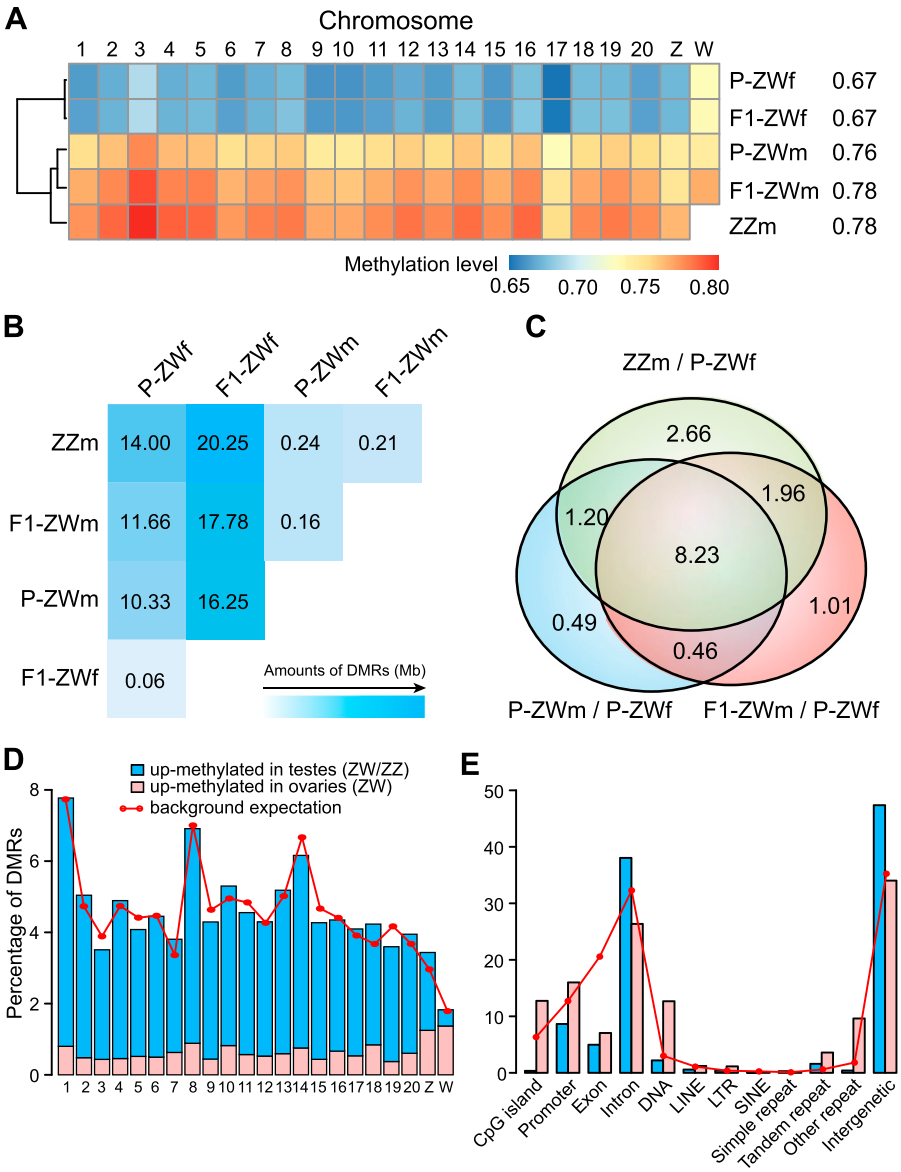

Figure 2. Genome-wide methylation level comparisons. (*A*) Methylation levels of different chromosomes. Numbers after the sample names represent methylation levels of the whole genome. Only CpGs with ≥10× coverage were used for analysis. (*B*) Total length (Mb) of DMRs identified in each pairwise comparison. (*C*) Venn diagrams for DMRs of P-ZWm/P-ZWf, F1-ZWm/P-ZWf, and ZZm/P-ZWf. Numbers represent the total length (Mb) of shared DMRs. (*D*) Percentage of testis/ovary DMRs on different chromosomes. Background expectation for each chromosome was calculated as the covered length (≥6× in all five samples) of each chromosome divided by the total covered length of all chromosomes. (*E*) Percentage of testis/ovary DMRs on different genomic elements. Background expectation for each element was calculated as the covered length of each element divided by the total covered length of the genome.

concordant within all three testis samples but different from the two ovary samples (see Methods). In total, we obtained ~18 Mb DMRs between testes and ovaries. Overall, these DMRs were widely distributed along the genome (Fig. 2D,E). However, in contrast to the overall up-methylated (i.e., the methylation level of a region in one sample is higher than that in the compared sample) bias in testis samples, both sex chromosomes had relatively higher amounts of DMRs up-methylated in the ovary samples, especially for the W chromosome (Fig. 2D). This pattern appears to be related to transposable element (TE) activity, as W and Z chromosomes were shown to have high and moderate percentages of TE components, respectively (Supplemental Fig. S4). Additionally, ovary-up DMRs were relatively enriched in CGIs, promoters, and repeated elements compared with testis-up DMRs, whereas testis-up DMRs tended to target introns and intergenic regions (Fig. 2E) (separate views for Z and W chromosomes are presented in Supplemental Fig. S6A,B).

We identified 2986 genes that harbored testis/ovary DMRs on their promoter regions, which we henceforth referred to as differentially methylated genes (DMGs). Gene ontology (GO) enrichment analysis (see Supplemental Methods) showed that DMGs that are up-methylated in ovaries were overrepresented in biological processes of development and morphogenesis, including reproductive structure development, female gonad development,

and oogenesis and spermatogenesis (Supplemental Table S6). In contrast, DMGs that are up-methylated in testes were enriched not only in developmental processes but also in the biological processes of responding to stimulus (e.g., cellular response to steroid hormone stimulus), signal transduction (e.g., steroid hormone mediated signaling pathway), and biological regulations (Supplemental Table S7). Taken together, these results imply that DNA methylation is involved in gonadal differentiation through regulating a series of biological processes.

We then identified DMRs where methylation levels were concordant within P-ZWm and F1-ZWm but different from ZZm, and obtained ~155 kb DMRs and 71 DMGs, which were relatively enriched in chromosome 9 and Z (Supplemental Fig. S6C,D), nevertheless without significant functional enrichment. We also compared the two pseudomale testis samples (P-ZWm + F1-ZWm) with three other samples (ZZm + P-ZWf + F1-ZWf) to identify regions with pseudomale specific methylation or demethylation after sex reversal; however, no solid DMRs were identified.

The sex determination pathway is the target of DNA methylation regulation

To gain insight into the role of epigenetic regulation in the sex determination pathway, we investigated the methylation status of 58 genomic loci documented to be associated with sex determination in other vertebrates. We observed a remarkable methylation contrast on the sex determination pathway between ZW female versus pseudomale and normal male gonads. In total, 16 of 58 (28%) sex-determination-related genes in tongue sole displayed significantly differential methylation patterns between testes and ovaries (Fig. 3A; Supplemental Table S8), which contrasts to ~14% (2986/21,516) over the whole genome (Fisher's exact test: $P = 0.0063$).

One of the DMGs, *dmrt1*, is a well-described and important gene required for male sexual development in a wide range of invertebrate (e.g., flies and worms) and vertebrate (e.g., birds and human) species (Raymond et al. 2000; Matson and Zarkower 2012). We recently reported that the tongue sole *dmrt1* was Z-linked, displayed testis-specific high expression during the critical stage of gonadal differentiation, and harbored a DMR in its promoter region which was up-methylated in ovaries compared with ZW/ZZ testes (Chen et al. 2014). To further examine when the differential methylation patterns emerge during gonad development, we quantified both the DNA methylation levels and the expression levels of this gene in gonad samples ranging from 4 d to 2 yr of age (see Methods). We found that in male gonads, this gene maintained low methylation levels throughout life, whereas in female gonads it underwent increasingly high levels of methylation at the start of the critical sex determination stage (Fig. 3B). Furthermore, its expression was repressed once the methylation had increased during this period. These results raise the possibility that *dmrt1* is the critical gene that responds to environmental change and triggers the sex reversal cascade in tongue sole.

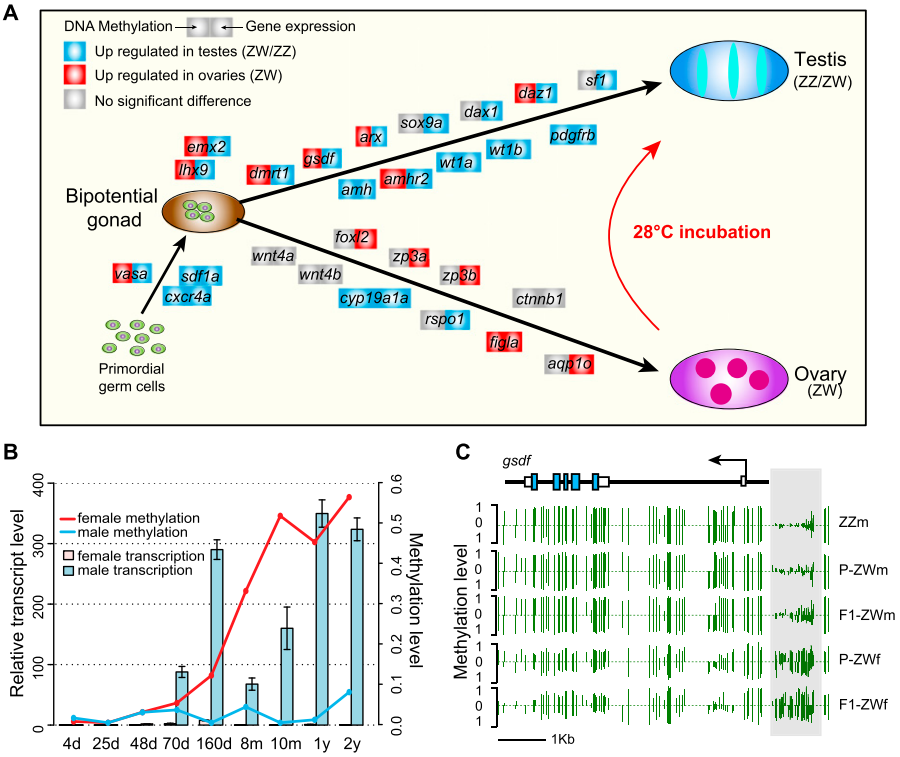

Figure 3. Differential methylation and sex determination. (*A*) Differentially methylated and differentially expressed genes in the putative sex determination pathway of tongue sole. For each gene presented in the pathway, the methylation (*left* square) or expression (*right* square) changes when comparing testes with ovaries are shown by different colors. (*B*) DNA methylation and transcription of *dmrt1* in different developmental stages after hatching. The methylation levels of different stages were estimated using bisulfite-PCR followed by TA-cloning with a pair of primers targeting the first exon, always using at least 10 randomly selected clones for sequencing for each stage. (*C*) DNA methylation profiles of *gsdf* in the five gonadal samples. Green vertical lines indicate the methylation level of cytosines. The light gray box indicates the DMR upstream of *gsdf*. Profiles of other DMGs in the pathway are presented in Supplemental Figure S8.

In addition, we observed a number of other DMGs that are reported to interact with each other in the sex-determining cascade or display conserved methylation patterns among vertebrates. *gsdf* (*gonadal somatic cell derived factor*) was recently reported to be the downstream gene of *dmY/dmrt1Y* (the Y copy that arose from duplication of the autosomal *dmrt1*) in the sex-determining cascade of medaka fish (Shibata et al. 2010; Myosho et al. 2012). RT-PCR analysis revealed that the tongue sole *gsdf* showed a highly similar expression pattern with *dmrt1* during sex differentiation (Supplemental Fig. S7; Supplemental Methods), and its promoter region was up-methylated in ovaries compared with ZW/ZZ testes (Fig. 3C), which might correspond to its approximately 23-fold up-regulated expression in testes compared with ovaries. *amh* (*anti-Mullerian hormone*) is another well-documented gene that mediates male sexual differentiation, whose expression is consistently higher in males than in females during sex differentiation in many vertebrates (Josso et al. 2005; Wu et al. 2010). The tongue sole *amh* showed a similar expression pattern (Supplemental Table S8) and harbored a DMR on the promoter region, which was up-methylated in ZW/ZZ testes in comparison with ovaries (Supplemental Fig. S8). Interestingly, we observed that *amhr2*, the receptor of *amh* (Wu et al. 2010), and *wt1a* and *wt1b*, potential activators of *amhr2* (Klattig et al. 2007), also showed differential methylation between ZW/ZZ testes and ovaries (Supplemental Fig. S8). Notably, consistent with the observation in sea bass (Navarro-Martín et al. 2011), the promoter of *cyp19a1a* was up-methylated in ZW/ZZ testes compared with ovaries (Supplemental Fig. S8), indicating the conserved regulatory role of DNA methylation on this gene.

Local dosage compensation of the Z chromosome in pseudomale testes

A critical question for animals with both GSD and ESD is how expression of genes in sex chromosomes is balanced after phenotypic sex reversal. We therefore investigated whether there is a specific dosage compensation mechanism on the Z chromosome to balance gene expression in pseudomales by calculating the ZZ testis to ZW testis ratio of gene expression (see Methods). The expected ratio for the complete absence of dosage compensation is two and for the presence of full compensation is one. We found that the expression levels of the Z chromosomal genes (Z-genes) in ZZ testes were on average 1.76 times higher than those in ZW testes (on average 1.73 for P-ZWm and 1.79 for F1-ZWm), a significantly higher ratio than the average male to female expression ratio of around 1.32 (Student's *t*-test: $P < 10^{-12}$ for P-ZW and $P < 10^{-9}$ for F1-ZW testis) estimated from the normal male and female whole fish samples (excluding gonads) (Fig. 4A; Supplemental Fig. S9). We then mapped the expression ratios along the Z chromosome and observed that a large proportion of the Z chromosome, from 0–12 Mb and 16–20 Mb (covering ~73% of the entire Z chromosome), pervasively displayed expression ratios of around two in ZZ testes versus ZW testes (Fig. 4A; Supplemental Fig. S9). This is different from the uniform distribution of dosage-compensated genes across the Z chromosome in female versus male whole fish, indicating a lack of dosage compensation for most of the Z chromosome in pseudomale testes.

Nevertheless, 162 genes, consisting of 17% (162/926) of the Z genes, showed different degrees of dosage compensation in ZW testes (Supplemental Table S9). Of note, there is one region (from

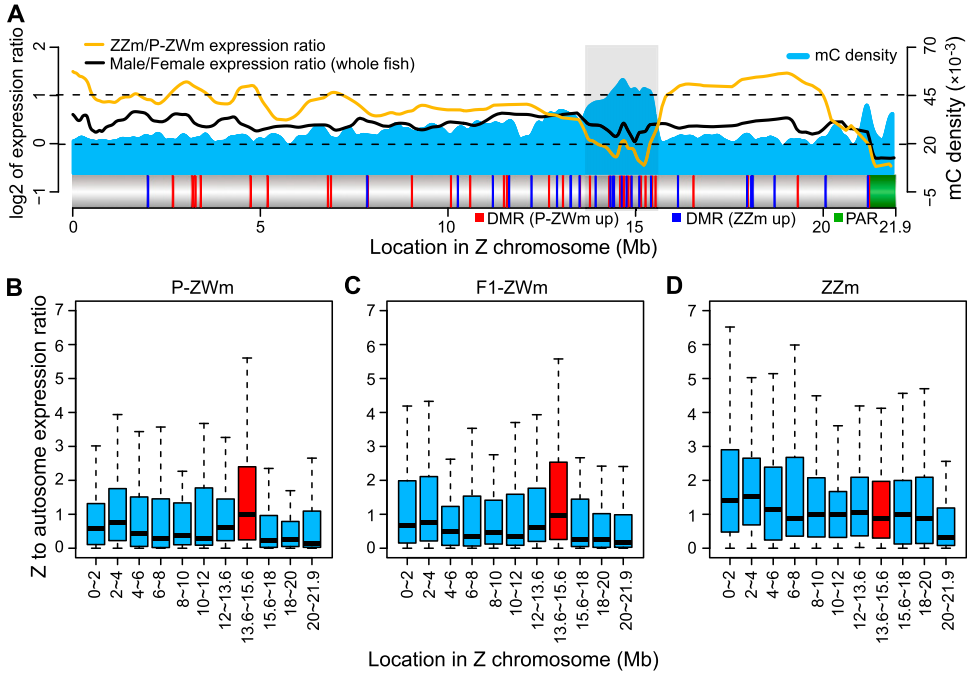

Figure 4. Dosage compensation of the Z chromosome in pseudomale testes. (*A*) Methylated cytosine (mC) density (5-kb window), log2-transformed expression ratios (running averages of 20 genes), and DMR profiles of the Z chromosome. The light gray box indicates the outstanding dosage-compensated region where DMRs were concentrated ([red vertical lines] DMRs that were up-methylated in P-ZWm, [blue vertical lines] DMRs that were up-methylated in ZZm), and the green block indicates the pseudoautosomal region (PAR) where Z and W chromosomes still pair in meiosis. Only 22 genes were annotated in PAR. Z-chromosomal to autosomal gene expression ratios (Z:A) in P-ZWm (*B*), F1-ZWm (*C*), and ZZm (*D*). The dosage compensation region (light gray box in *A*) is plotted in red. For each Z interval, the expression level of each Z-gene was first divided by the median expression level of all autosomal genes, then the Z:A ratios in each interval were plotted.

13.6 to 15.6 Mb) that is specifically enriched with dosage-compensated genes in pseudomale testes (Fig. 4A). To answer whether the dosage compensation in this region was the result of down-regulating ZZ testis genes or up-regulating ZW testis genes, we calculated the expression ratio of each Z-gene relative to the median of all autosome genes (Z:A) in ZW testes and ZZ testes separately (see Methods). In ZW testes, the Z:A ratios in this region were significantly higher than those in any other region of the Z chromosome (Mann-Whitney *U*-test: $P < 10^{-7}$ for P-ZWm and $P < 10^{-6}$ for F1-ZWm), displaying a median value close to one and close to most Z regions in ZZ testes (Fig. 4B–D). This indicated that dosage compensation of this region in ZW testes is due to general up-regulation of gene expression. Interestingly, we also found that this dosage-compensated region exhibited a very high density of methylated cytosines (Fig. 4A; Supplemental Fig. S9) and was enriched with Z-chromosomal DMRs between pseudomale and normal male (harbored ~38% of the DMRs on the entire Z chromosome). These results suggest that DNA methylation might play an important role in regulating dosage compensation at the 13.6- to 15.6-Mb region of the Z chromosome.

The unique dosage compensation region found in pseudomale testes implies the potentially critical role of genes in this region with regard to male development, particularly testis development. As expected, we found that some genes from this region were involved in spermatogenesis. For example, *piwil2* (*piwi-like 2*), a member of the *piwi* family of genes that exhibit conserved functions relating to transposon silencing during spermatogenesis (Reuter et al. 2009; Wang et al. 2009; De Fazio et al. 2011), displayed a moderate degree of dosage compensation in both ZW testes with ZZ to ZW testis expression ratios of 1.31 and 1.70 in P-ZWm and F1-ZWm, respectively, but no compensation was observed in ovaries (expression ratio of ZZ testis/ZW ovary is 3.69) (Supplemental Table S9). Another example is *pik3r1* (*phosphoinositide-3-kinase, regulatory subunit 1* [*p85 alpha*]), a required component in androgen-stimulated PI3K/Akt pathway activation (Castoria et al. 2003; Sun et al. 2003), which plays a central role in the self-renewal division of spermatogonial stem cells (Lee et al. 2007). This gene displayed a high degree of dosage compensation, with ZZ to ZW testis expression ratios of 0.61 and 0.79 in P-ZWm and F1-ZWm, respectively (Supplemental Table S9), whereas highly suppressed expression of this gene was observed in ovaries (expression ratio of ZZ testis/ZW ovary is 8.61). More examples are listed in Supplemental Table S9.

Suppression of female-specific gene expression in pseudomales

We next investigated the expression of W-chromosomal genes (W-genes) in ovaries and ZW testes and found that among the 317 annotated protein-coding genes on the W chromosome, up to 120 genes (38%) were silenced or had very low expression level (RPKM <1) in both ovaries and ZW testes. Surprisingly, however, up to 94% (185/197) of the 197 remaining genes were still actively transcribed (RPKM ≥1) in either P-ZWm or F1-ZWm. Furthermore, among the 94 genes that were differentially expressed (twofold RPKM change) between ovaries and ZW testes, 46% (43/94) were up-regulated in both ZW testes relative to ovaries (Fig. 5A). Because a large proportion of W-genes are still active in pseudomales, many of the W-genes thus appear to be harmless for male phenotypic development. This may be explained by the young age of the ZW chromosome system in tongue sole, which evolved only around 30 Mya (Chen et al. 2014). We found that for all 317 annotated protein-coding genes on the W chromosome, the paralogs of 272 genes (86%) can be detected on the Z chromosome with high identity (94% in median) (Fig. 5B). This implies that most of the W genes are still too young to develop neofunction in comparison to their counterparts on the Z chromosome; thus, they are generally harmless for male development. Furthermore, their expression may even compensate the gene dosage of their Z counterparts. This is supported by the observation that the expression sum of Z-W paralogous genes in pseudomale testes was close to the dosage of Z-genes in normal males (Fig. 5C; Supplemental Fig. S10).

Despite the relatively young age of the ZW sex determination system in tongue sole, there are three W-specific genes that do not share any functional homologs on the Z chromosome or autosomes (Chen et al. 2014). Although two of these genes were silenced both in ovaries and ZW testes, the third gene, *figla* (*factor in the germline alpha*), which is a germ cell-specific basic helix-loop-helix (bHLH) factor required for ovarian follicle formation (Liang et al. 1997; Soyal et al. 2000; Bayne et al. 2004; Kanamori et al. 2008), displayed interesting expression patterns. RNA-seq data suggested that it could produce two different transcripts by targeting mutually exclusive first exons (Fig. 5D). We found that the transcript with the alternative first exon AFE1 was specifically transcribed in ovaries (Fig. 5E) and was the canonical splicing form that is expressed and is functional in other vertebrate species (Liang et al. 1997; Kanamori et al. 2008). However, the pseudomale testes lack this splicing form, and only express the other transcript with AFE2 (Fig. 5F; see Supplemental Methods). Domain annotation suggested that only the ovary-specific splicing form contains the helix-loop-helix DNA binding domain that is critical to the binding with other proteins (Fig. 5C), so that ovarian follicles can be formed (Liang et al. 1997; Chaudhary and Skinner 1999). It is interesting that the use of different splicing forms by ovaries and pseudomale testes is strongly associated with DNA methylation. There is a DMR between ovaries and pseudomale testes around the AFE1 whose methylation pattern is consistent with the operation of alternative splicing, as this region has a high level of methylation in ovaries (recruiting this exon) but hardly any methylation in pseudomale testes (recruiting another exon) (Fig. 5D). It thus appears that this female-bias gene has been suppressed by alternative splicing via DNA methylation regulation, leading to function loss in pseudomales.

Discussion

Roles of epigenetic regulation in sex determination and sex reversal

In this study, we used half-smooth tongue sole as a model to characterize and compare the epigenetic and transcription pattern of sex-reversed individuals and normal individuals, and to provide insight into the molecular relationship between GSD and ESD. To our knowledge, our study provides the first comprehensive investigation of gene expression and regulation at the whole genome level in a GSD species that exhibits the ESD phenomenon.

Studies on model organisms have revealed that sex development is regulated by a network of genes that have simultaneous input on downstream cascades (Angelopoulou et al. 2012; Munger and Capel 2012). These key master genes in the network are evolutionarily highly conserved and expressed plastically across species (Angelopoulou et al. 2012; Munger and Capel 2012). It has been hypothesized that the functional conserved genetic components involved in gonadal differentiation in GSD animals can also be instrumental in the expression of ESD (Crews and Bull 2009).

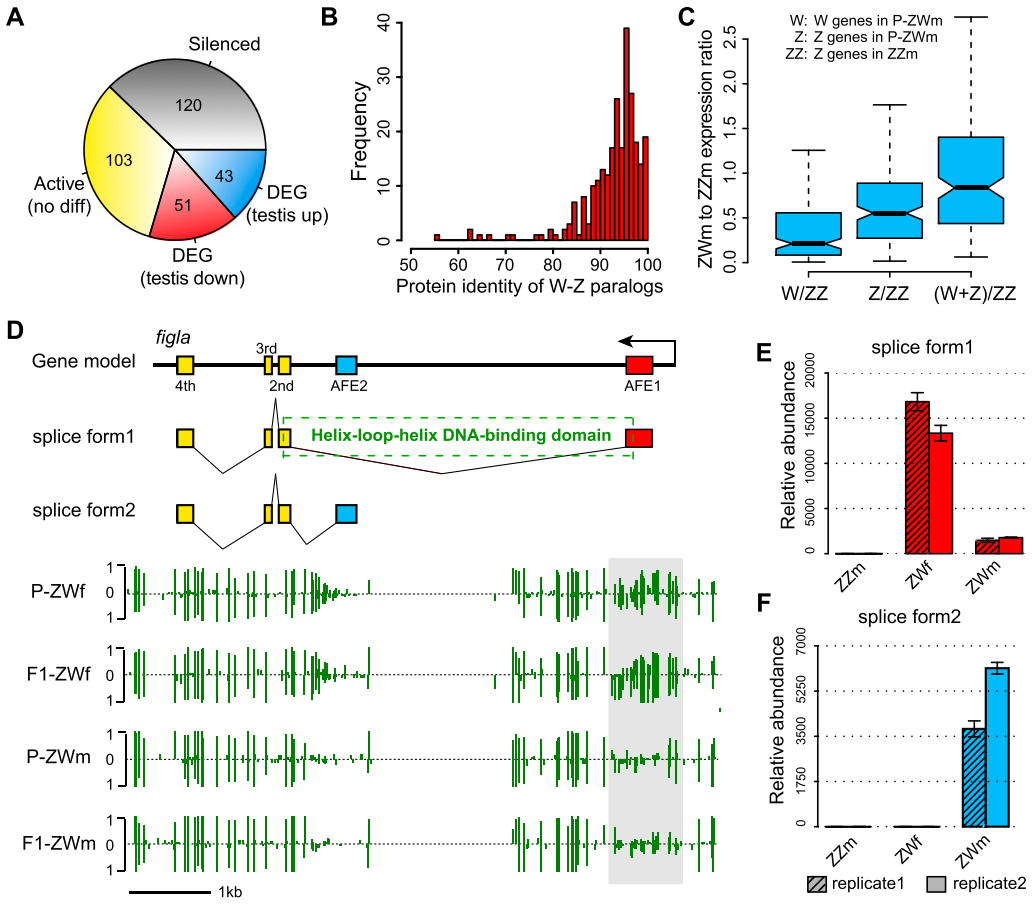

Figure 5. W-genes expression pattern in pseudomale testes. (*A*) Categories of W-genes based on gene expression levels in ovaries and pseudomale testes. (*B*) Frequency distribution of protein identities for the 272 W-Z paralogous pairs. (*C*) Pseudomale to normal-male expression ratios calculated from the 272 W-genes and their Z-counterparts. From *left* to *right*: The first box represents expression ratios of P-ZWm W-genes to ZZm Z-genes; the second box represents expression ratios of P-ZWm Z-genes to ZZm Z-genes; and the third box represents expression ratios of the sum of P-ZWm W- and Z-genes to ZZm Z-genes, indicating that the expression sum of W-Z paralogous genes in pseudomale testes was close to the dosage of Z-genes in normal males. (*D*) Alternative splicing and methylation profile of *figla*. The *upper* section indicates the gene model and the two splice forms of *figla*, with the dashed box indicating the position of the basic helix–loop–helix domain, which is coded by first and second exons of splice form1. The *lower* section shows the DNA methylation profiles in ovaries and pseudomale testes, with the light gray box indicating the position of DMR and green vertical lines indicating methylation levels of cytosines. (*E,F*) Expression of the two splice forms of *figla* in testes of normal and pseudomales and female ovaries as determined by RT-PCR. (AFE) Alternative first exon.

However, due to the limited genetic resources available for ESD animals, previous analysis has been limited to a few genes. Our results show that a majority of the genes that are involved in sex determination in other vertebrates are highly conserved in the tongue sole genome. Furthermore, their orthologs display substantial differences in gonadal expression or methylation regulation in tongue sole, suggesting a conserved evolutionary core of the sex determination toolkit. We also show that the differences in gonadal expression and methylation pattern of these key genes are consistent between females and the pseudomales that develop from sex-reversed females as a result of environmental influence. This result pinpoints how the master regulatory genes in the sex determination network used in GSD are also utilized during the ESD process to trigger the cascade of alternate gonad development.

It has long been proposed that epigenetic regulation may play an important role in ESD species in response to environmental influences, leading to the gonad developmental change (Manolakou et al. 2006). The *cyp19a1a* is one of the first cases to show an epigenetic mechanism that mediates temperature effects in sexual reversal (Navarro-Martín et al. 2011). We demonstrate here that regulation by DNA methylation in sexual reversal of tongue sole is not limited to this gene but is applied across the entire sex determination network, which shows an enrichment of DMGs (Fig. 3A). Due to limits in the tissue size required for large-scale sequencing, as well as the challenge of distinguishing the phenotypic sex status of the premature gonad during the sensitive development stage, it is impossible to compare the gene expression and methylation between pre- and post-sexual reversal; thus it is difficult to conclude whether the methylation difference is the cause or consequence of sex reversal. To do so would require more detailed functional experiments. However, our current results provide some insight. For instance, our expression profiling reveals that the first detectable high expression of the *dmrt1* gene starts during the period of gonadal sex determination in tongue sole at ~70 d after hatching and persists at high levels in testes once sex has become differentiated (Fig. 3B). Interestingly, the male specific expression pattern seems to be implemented by the hyper-

methylation in ovaries on its promoter region. Moreover, this hypermethylation pattern can be eliminated in genotypic females by high temperature incubation and prevent transcript silencing during the critical developmental stage, leading the development of undifferentiated gonad into testes. Of note, medaka also exhibit reduced expression of *dmY/dmrt1Y w*hen XY males change into females (Matsuda et al. 2002). It thus appears that ESD can override the standard GSD expression pathways through epigenetic regulation of just a few core genes in the sex determination network, with these affected genes inducing and cascading downstream effects that mediate the expression of sex-specific morphological and physiological traits. As shown in Figure 3, the transcription of several key genes in ovary development is suppressed, whereas the activity of testis development related genes is reactivated after the sexual reversal from genotypic female to phenotypic males by the regulation of DNA methylation.

The role of DNA methylation in mediating gene expression is well documented. It has been proposed recently that the regulation of gene activity by DNA methylation can also be achieved by control of transcript splicing (Shukla et al. 2011). Here, we provide an interesting example, showing that the transcript splicing of a female specific gene *figla* is associated with DNA methylation. The specific role of methylation in alternative splicing will have to be determined by further experiments, but the case presented here and its important function may shed light on the understanding of this regulatory mechanism.

Resolution of sex-linked gene dosage inequality during sex reversal

Dosage compensation is a process that normalizes the expression level of genes from unequal copy numbers of sex chromosomes in GSD species (Mank 2009, 2013). By comparison of gene expression in females versus males, our previous study suggests that the tongue sole has incomplete dosage compensation in its Z chromosome (Chen et al. 2014). The phenomenon of sexual reversal in tongue sole raises the similar scenario of Z chromosome dosage during the sex reversal from genotypic females to phenotypic males since they have only one copy of Z relative to the normal males. Our result suggests that low dosage compensation on the Z chromosome in the pseudomale is sufficient to achieve normal testis function after sex reversal. Alternatively, the insufficient dosage of Z genes in pseudomales may have been compensated by their undifferentiated paralogs on the W chromosome. Therefore, the dosage compensation is not necessarily fully developed in this differentiating young ZW system, especially when the two sex chromosomes have not been accumulating too many sex-linked mutations and sexually antagonistic genes.

There is one region standing out from the Z chromosome and clustering with a significantly high number of genes depending on dosage compensation in testes (Fig. 4). Consistent with the above hypothesis about dosage compensation, this region is extremely diverged from the W chromosome, containing a significantly low number of paralogous genes in W (19.7%, whole Z chromosome is 31.7%; $P = 0.01$). This region also bears a high density of methylated cytosine and is significantly hypermethylated (Fig. 4A). It is worth noting that a similar dosage-compensated region has been found on the chicken Z chromosome, where a majority of compensated genes is also located within the so-called male hypermethylated region (MHM) (Teranishi et al. 2001; Melamed and Arnold 2007). It has been suspected that dosage compensation in chicken could be mediated by noncoding RNA in MHM, which recruits a histone acetyltransferase to stimulate up-regulation of genes in that region (Teranishi et al. 2001; Melamed and Arnold 2007). A similar scenario may also apply in pseudomales of tongue sole but through a different trajectory of DNA methylation. Further experiments are needed to elucidate the exact mechanisms. In spite of the similar mechanisms of dosage compensation in chicken and tongue sole, we found no significantly homologous relationships between these two regions, suggesting that they evolved independently.

Another potential dosage dilemma in the combined GSD and ESD system of tongue sole fish concerns the W chromosome of pseudomales, as pseudomales would need to develop a strategy to inhibit the female-beneficial/male-detrimental genes, such as the W-linked *figla*, during the development of a functional testis instead of an ovary. The tongue sole *figla* provides an excellent example as to how sex reversed individuals overcome this contradiction. Instead of being completely silenced, this gene is suppressed by an alternative splicing mechanism and associated with DNA methylation regulation. Meanwhile, we also note that only a fraction of the W genes have been repressed in pseudomales, due to the ongoing diverging status of this recently evolved neo-ZW system. We therefore propose that this incompletely differentiated ZW chromosome system, in which neither of the sex chromosomes accumulates sex-specific antagonistic genes, together with epigenetic regulatory mechanisms, may be important in creating the plasticity to develop ESD within a GSD system.

Transgenerational epigenetic inheritance of sex reversal

Transgenerational epigenetic inheritance is the epigenetic effect on the phenotype that can be passed down to subsequent generations and cannot be explained by the changes in the primary DNA sequence and Mendelian genetics (Youngson and Whitelaw 2008; Daxinger and Whitelaw 2010, 2012). It has only rarely been reported in mammals due to the epigenetic reprogramming between generations in mammals that erases and resets most of the epigenetic marks (Daxinger and Whitelaw 2010, 2012; Feng et al. 2010). However, the erasing of epigenetic marks is not a universal rule in animals. For example, DNA methylation is not reset in early development in zebrafish (Macleod et al. 1999), pinpointing the possibility of transgenerational epigenetic inheritance in fishes. This has been confirmed by a recent study in zebrafish, which revealed the heredity of parental DNA methylatome in offspring embryo (Jiang et al. 2013; Potok et al. 2013). A unique advantage of half-smooth tongue sole is that the offspring of pseudomales can spontaneously develop into functional pseudomales without environmental stimuli. An explanation for this is that novel DNA methylation patterns are established in bipotential germ cells when the juvenile fish incubate at high temperature and alter the developmental fate of the germ cells to the opposite sex. These new methylation patterns are then imprinted in the genome and not erased and reset during the development of new generations, so they appear in the offspring. This is supported by our observation that the methylation differences between parental pseudomales versus females are highly consistent with the differences between offspring pseudomales versus females (Fig. 2C; Supplemental Fig. S5).

However, since it is only the methylation patterns of paternal alleles in offspring originating from their pseudomale fathers that are rewritten, while their maternal alleles, which have not been subjected to high temperature exposure, do not contain the novel methylation marks, the following question arises: Why (or how) does the monoallelic methylation change in offspring promote

sexual reversal, leading to most or all of the genotypic females' offspring developing into males? Our previous population survey for a chromosomal inheritance pattern revealed that most or all ZW offspring of the pseudomale inherit their Z chromosomes exclusively from their fathers, and no maternal Z chromosome has been observed in ZW offspring (Chen et al. 2014). This can be explained by the fact that during the spermatogenesis process, no germ cell with only the W chromosome produced by pseudomales can develop to sperm and survive (Chen et al. 2014). Therefore, the ZW offspring of pseudomales only inherit the paternal Z allele, of which the methylation patterns have been altered by environmental stimulation and switch off the ovary developmental pathways. Once the novel methylation marks are inherited, the offspring, which do not have epigenetic reprogramming mechanisms in early developmental stages to reset the epigenetic marks, maintain their paternal methylation pattern through their lifetime, thus directly develop into functional males. Finally, the stable inheritance pattern of the Z chromosome in the pseudomale family offers the possibility of consistent *trans*-generational epigenetic inheritance of sexual reversal phenomenon in tongue sole fish (Supplemental Fig. S11).

Methods

Sample collection

Normal males (ZZm), parental females (P-ZWf), and induced pseudomales (P-ZWm) were collected from Laizhou Mingbo, Co., from 2008 to 2010 (Fig. 1C). One of the pseudomales (P-ZWm) was crossed with one of the females (P-ZWf) to produce the next generation of females (F1-ZWf) and pseudomales (F1-ZWm), which were collected as F1 samples when they were mature (Fig. 1C). For each of the five samples, two biological replicates were utilized, with each replicate being pooled by five fish. The phenotype and genotype of each selected fish were identified by the histological analysis and PCR validation using the W chromosome specific marker (Chen et al. 2012).

DNA isolation, BS-seq library construction, and sequencing

DNA were isolated from five pooled gonads of the same replicate, then 5 μg DNA was used to perform the bisulfite conversion and BS-seq. The bisulfite conversion of sample DNA was carried out using a modified NH_4HSO_3-based protocol (Hayatsu et al. 2006). The paired-end library construction and sequencing were carried out using Illumina HiSeq 2000, according to the manufacturer's instructions (Illumina). We also mixed 25 ng cl857 Sam7 Lambda DNA in each sample to use as a conversion quality control for each library.

BS-seq analysis

The Lambda genome was merged with the reference genome of *Cynoglossus semilaevis* so that reads originating from the unmethylated control DNA could be aligned. BS-seq reads were mapped to the reference genome using SOAP2 (Li et al. 2009) as described in Bonasio et al. (2012), allowing up to six mismatches for 90-bp paired-end reads. Multiple reads mapping to the same position were regarded as PCR duplicates, and only one of them was kept. Bases with a quality score <20 were not considered for subsequent analysis.

The error rate of each library (sum of the nonconversion rate and T/C sequencing errors) was calculated as the total number of sequenced Cs divided by the total sequencing depth for sites corresponding to Cs in the Lambda genome. The error rate for each library was ~0.5% (Supplemental Table S2). To distinguish true mCs from false positives, we used a model based on the binomial distribution B(n,p) following Bonasio et al. (2012), and only the mCs with FDR (Benjamini et al. 2001) adjusted *P*-values < 0.01 were considered true positives.

Methylation level calculation

The methylation level of an individual CpG was determined by the number of reads containing a C at the site of interest divided by the total number of reads covering the site. The methylation level of a specific region was determined by the sum of methylation levels of individual CpGs in the region divided by the total number of covered CpGs in this region.

Differential methylation analysis

Two-way ANOVA was used to identify DMRs between two groups of samples using a 500-bp sliding window with the step length of 250 bp. To ensure the power of statistical tests, only windows with at least six (three per strand) informative CpGs (≥3× coverage) in all replicates of the two compared samples were considered. The two independent variables for ANOVA were sample and cytosine position. For each window, we first calculated the variance between samples and the variance between two replicates of each sample (variance caused by random error), then used F-test to calculate the *P*-value of each window by comparing the sample variance and random error variance. *P*-values were then adjusted by FDR (Benjamini et al. 2001). Only windows with adjusted *P*-value < 0.05 and greater than twofold methylation level change were considered as DMRs. Differentially methylated genes (DMGs) were defined as genes containing DMRs in their putative promoter regions (TSS −2 kb to +500 bp).

Fisher's exact test and χ^2 test (Beißbarth and Speed 2004) were employed to estimate whether the DMGs were enriched in specific GO categories when compared with background genes. *P*-values were adjusted by FDR (Benjamini et al. 2001), and the adjusted *P*-value < 0.05 was chosen as the significant threshold (see Supplemental Methods for details).

To track the DNA methylation status of *dmrt1* during gonadal development, we performed bisulfite-PCR for genomic DNA extracted from genetic female and male gonads at different developmental stages using a pair of primers targeting the first exon of *dmrt1*. PCR products were subjected to TA-cloning, and at least 10 clones were randomly selected for Sanger sequencing for each stage (see Supplemental Methods for details).

Gene expression analysis

Gonadal transcriptome data of P-ZWf, P-ZWm, F1-ZWm, and ZZm, as well as transcriptome data of whole fish (excluding gonads) of female and normal male, were from the tongue sole genome project (Chen et al. 2014), and these transcriptome data have been deposited in the NCBI Sequence Read Archive (ZZm: SRX106096; P-ZWf: SRX106097; P-ZWm: SRX106098; F1-ZWm: SRX106099; female whole fish: SRX106100; male whole fish: SRX106103). TopHat v2.0.4 package (Trapnell et al. 2009) was used to map transcriptome reads to the tongue sole genome, with parameters of -a/–min-anchor 8, -m/–splice-mismatches 0, -i/–min-intron-length 50, -I/–max-intron-length 500000,–segment-mismatches 3 and–segment-length 25 for 75-bp reads and 30 for 90-bp reads. Gene expression levels were measured by RPKM (reads per kilobase of gene per million mapped reads) (Mortazavi et al. 2008) and adjusted by a scaling normalization method (Robinson and Oshlack 2010).

Expression of *gsdf* in different developmental stages and expression of *figla* in different adult gonads were quantified by conventional RT-PCR (see Supplemental Methods for details). RT-PCR analysis results of *dmrt1* at different developmental stages were obtained from Chen et al. (2014) by adding the data of two additional time points (8 m and 10 m) using the 1 y sample as a control (Chen et al. 2014).

Dosage compensation analysis of Z chromosomal genes

The "male to female" or "male to pseudomale" gene expression ratios were used to measure the degree of dosage compensation for each Z-gene in female (ZW) or pseudomale (ZW) relative to normal male (ZZ), calculated as the RPKM ratio of each Z-gene between two compared samples. Only genes with RPKM greater than one in both compared samples were considered, and genes with "male to female" or "male to pseudomale" RPKM ratios between 1/1.5 and 1.5 were defined as dosage-compensated genes. In Figure 4A and Supplemental Figure S9, RPKM ratios were log2-transformed, and the running averages of 20 genes were plotted across the Z chromosome to visualize the chromosome-wide dosage compensation pattern.

To investigate whether dosage compensation of the 13.6- to15.6-Mb region of the Z chromosome in pseudomale testes was the result of down-regulating male expressions or up-regulating pseudomale expressions, we calculated the expression ratio of each Z-gene to the median of autosomal genes (Z:A) in pseudomale testes and normal male testes separately, then Z:A ratios of the Z-genes in each ~2 Mb interval were visualized by box plot (Fig. 4B–D). Only genes with RPKM greater than one in at least one of the six transcriptome samples were used for Z:A analysis.

Data access

DNA methylome data of the five gonadal samples from this study have been submitted to the NCBI Gene Expression Omnibus (GEO; http://www.ncbi.nlm.nih.gov/geo/) under accession number GSE41129.

Acknowledgments

We thank Jacobus J. Boomsma and M. Thomas P. Gilbert for their critical comments on this manuscript. This work was supported by grants from the National Nature Science Foundation of China (31130057, 41006107, and 31072202), State 863 High-Technology R&D Project of China (2012AA092203, 2012AA10A403-2), Special Fund for Agro-scientific Research in the Public Interest (200903046) of China, Taishan Scholar Project Fund of Shandong of China, JSPS RONPAKU (Dissertation PhD) Program, China National Genebank of China, and a Marie Curie International Incoming Fellowship (300837) to G.Z.

References

Angelopoulou R, Lavranos G, Manolakou P. 2012. Sex determination strategies in 2012: Towards a common regulatory model? *Reprod Biol Endocrinol* **10:** 13.

Bayne RAL, Martins da Silva SJ, Anderson RA. 2004. Increased expression of the FIGLA transcription factor is associated with primordial follicle formation in the human fetal ovary. *Mol Hum Reprod* **10:** 373–381.

Beißbarth T, Speed TP. 2004. GOstat: Find statistically overrepresented Gene Ontologies within a group of genes. *Bioinformatics* **20:** 1464–1465.

Benjamini Y, Drai D, Elmer G, Kafkafi N, Golani I. 2001. Controlling the false discovery rate in behavior genetics research. *Behav Brain Res* **125:** 279–284.

Bonasio R, Li Q, Lian J, Mutti NS, Jin L, Zhao H, Zhang P, Wen P, Xiang H, Ding Y et al. 2012. Genome-wide and caste-specific DNA methylomes of the ants *Camponotus floridanus* and *Harpegnathos saltator*. *Curr Biol* **22:** 1755–1764.

Bull JJ. 1983. *Evolution of sex determining mechanisms*. Benjamin/Cummings Publishing Company, Inc., Menlo Park, CA.

Castoria G, Lombardi M, Barone MV, Bilancio A, Di Domenico M, Bottero D, Vitale F, Migliaccio A, Auricchio F. 2003. Androgen-stimulated DNA synthesis and cytoskeletal changes in fibroblasts by a nontranscriptional receptor action. *J Cell Biol* **161:** 547–556.

Chaudhary J, Skinner MK. 1999. Basic helix-loop-helix proteins can act at the E-box within the serum response element of the *c-fos* promoter to influence hormone-induced promoter activation in Sertoli cells. *Methods Enzymol* **13:** 774–786.

Chen S, Tian Y, Yang J, Shao C, Ji X, Zhai J, Liao X, Zhuang Z, Su P, Xu J, et al. 2009. Artificial gynogenesis and sex determination in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar Biotechnol (NY)* **11:** 243–251.

Chen S, Ji X, Shao C, Li W, Yang J, Liang Z, Liao X, Xu G, Xu Y, Song W. 2012. Induction of mitogynogenetic diploids and identification of WW super-female using sex-specific SSR markers in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar Biotechnol (NY)* **14:** 120–128.

Chen S, Zhang G, Shao C, Huang Q, Liu G, Zhang P, Song W, An N, Chalopin D, Volff J, et al. 2014. Whole-genome sequence of a flatfish provides insights into ZW sex chromosome evolution and adaptation to a benthic lifestyle. *Nat Genet* doi: 10.1038/ng.2890.

Crews D, Bull JJ. 2009. Mode and tempo in environmental sex determination in vertebrates. *Semin Cell Dev Biol* **20:** 251–255.

Daxinger L, Whitelaw E. 2010. Transgenerational epigenetic inheritance: More questions than answers. *Genome Res* **20:** 1623–1628.

Daxinger L, Whitelaw E. 2012. Understanding transgenerational epigenetic inheritance via the gametes in mammals. *Nat Rev Genet* **13:** 153–162.

De Fazio S, Bartonicek N, Di Giacomo M, Abreu-Goodger C, Sankar A, Funaya C, Antony C, Moreira PN, Enright AJ, O'Carroll D. 2011. The endonuclease activity of Mili fuels piRNA amplification that silences LINE1 elements. *Nature* **480:** 259–263.

Devlin RH, Nagahama Y. 2002. Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences. *Aquaculture* **208:** 191–364.

Feng S, Jacobsen SE, Reik W. 2010. Epigenetic reprogramming in plant and animal development. *Science* **330:** 622–627.

Hayatsu H, Tsuji K, Negishi K. 2006. Does urea promote the bisulfite-mediated deamination of cytosine in DNA? Investigation aiming at speeding-up the procedure for DNA methylation analysis. *Nucleic Acids Symp Ser (Oxf)* **50:** 69–70.

Janzen FJ. 1995. Experimental evidence for the evolutionary significance of temperature dependent sex determination. *Evolution* **49:** 864–873.

Jiang L, Zhang J, Wang J-J, Wang L, Zhang L, Li G, Yang X, Ma X, Sun X, Cai J. 2013. Sperm, but not oocyte, DNA methylome is inherited by zebrafish early embryos. *Cell* **153:** 773–784.

Josso N, Belville C, di Clemente N, Picard JY. 2005. AMH and AMH receptor defects in persistent Müllerian duct syndrome. *Hum Reprod Update* **11:** 351–356.

Kanamori A, Toyama K, Kitagawa S, Kamehara A, Higuchi T, Kamachi Y, Kinoshita M, Hori H. 2008. Comparative genomics approach to the expression of *figα*, one of the earliest marker genes of oocyte differentiation in medaka (*Oryzias latipes*). *Gene* **423:** 180–187.

Kato Y, Kobayashi K, Watanabe H, Iguchi T. 2011. Environmental sex determination in the branchiopod crustacean *Daphnia magna*: Deep conservation of a *Doublesex* gene in the sex-determining pathway. *PLoS Genet* **7:** e1001345.

Klattig J, Sierig R, Kruspe D, Besenbeck B, Englert C. 2007. Wilms' tumor protein Wt1 is an activator of the anti-Müllerian hormone receptor gene *Amhr2*. *Mol Cell Biol* **27:** 4355–4364.

Lee J, Kanatsu-Shinohara M, Inoue K, Ogonuki N, Miki H, Toyokuni S, Kimura T, Nakano T, Ogura A, Shinohara T. 2007. Akt mediates self-renewal division of mouse spermatogonial stem cells. *Development* **134:** 1853–1859.

Li R, Yu C, Li Y, Lam TW, Yiu SM, Kristiansen K, Wang J. 2009. SOAP2: An improved ultrafast tool for short read alignment. *Bioinformatics* **25:** 1966–1967.

Liang L, Soyal SM, Dean J. 1997. FIGα, a germ cell specific transcription factor involved in the coordinate expression of the zona pellucida genes. *Development* **124:** 4939–4947.

Macleod D, Clark VH, Bird A. 1999. Absence of genome-wide changes in DNA methylation during development of the zebrafish. *Nat Genet* **23:** 139–140.

Mank JE. 2009. The W, X, Y and Z of sex-chromosome dosage compensation. *Trends Genet* **25:** 226–233.

Mank JE. 2013. Sex chromosome dosage compensation: Definitely not for everyone. *Trends Genet* **29:** 677–683.

Manolakou P, Lavranos G, Angelopoulou R. 2006. Molecular patterns of sex determination in the animal kingdom: A comparative study of the biology of reproduction. *Reprod Biol Endocrinol* **4:** 59.

Matson CK, Zarkower D. 2012. Sex and the singular DM domain: Insights into sexual regulation, evolution and plasticity. *Nat Rev Genet* **13:** 163–174.
Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, Kobayashi T, Morrey CE, Shibata N, Asakawa S, Shimizu N. 2002. *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* **417:** 559–563.
Matsumoto Y, Crews D. 2012. Molecular mechanisms of temperature-dependent sex determination in the context of ecological developmental biology. *Mol Cell Endocrinol* **354:** 103–110.
Melamed E, Arnold AP. 2007. Regional differences in dosage compensation on the chicken Z chromosome. *Genome Biol* **8:** R202.
Mitchell NJ, Janzen FJ. 2010. Temperature-dependent sex determination and contemporary climate change. *Sex Dev* **4:** 129–140.
Mortazavi A, Williams BA, McCue K, Schaeffer L, Wold B. 2008. Mapping and quantifying mammalian transcriptomes by RNA-Seq. *Nat Methods* **5:** 621–628.
Munger SC, Capel B. 2012. Sex and the circuitry: Progress toward a systems-level understanding of vertebrate sex determination. *Wiley Interdiscip Rev Syst Biol Med* **4:** 401–412.
Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, Fujiyama A, Naruse K, Hamaguchi S, Sakaizumi M. 2012. Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. *Genetics* **191:** 163–170.
Narita S, Kageyama D, Nomura M, Fukatsu T. 2007. Unexpected mechanism of symbiont-induced reversal of insect sex: Feminizing *Wolbachia* continuously acts on the butterfly *Eurema hecabe* during larval development. *Appl Environ Microbiol* **73:** 4332–4341.
Navarro-Martín L, Viñas J, Ribas L, Díaz N, Gutiérrez A, Di Croce L, Piferrer F. 2011. DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet* **7:** e1002447.
Ospina-Alvarez N, Piferrer F. 2008. Temperature-dependent sex determination in fish revisited: Prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS ONE* **3:** e2837.
Potok ME, Nix DA, Parnell TJ, Cairns BR. 2013. Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153:** 759–772.
Quinn AE, Georges A, Sarre SD, Guarino F, Ezaz T, Graves JA. 2007. Temperature sex reversal implies sex gene dosage in a reptile. *Science* **316:** 411.
Raymond CS, Murphy MW, O'Sullivan MG, Bardwell VJ, Zarkower D. 2000. *Dmrt1*, a gene related to worm and fly sexual regulators, is required for mammalian testis differentiation. *Genes Dev* **14:** 2587–2595.
Reuter M, Chuma S, Tanaka T, Franz T, Stark A, Pillai RS. 2009. Loss of the Mili-interacting Tudor domain-containing protein-1 activates transposons and alters the Mili-associated small RNA profile. *Nat Struct Mol Biol* **16:** 639–646.
Robinson MD, Oshlack A. 2010. A scaling normalization method for differential expression analysis of RNA-seq data. *Genome Biol* **11:** R25.
Shibata Y, Paul-Prasanth B, Suzuki A, Usami T, Nakamoto M, Matsuda M, Nagahama Y. 2010. Expression of gonadal soma derived factor (GSDF) is spatially and temporally correlated with early testicular differentiation in medaka. *Gene Expr Patterns* **10:** 283–289.
Shukla S, Kavak E, Gregory M, Imashimizu M, Shutinoski B, Kashlev M, Oberdoerffer P, Sandberg R, Oberdoerffer S. 2011. CTCF-promoted RNA polymerase II pausing links DNA methylation to splicing. *Nature* **479:** 74–79.
Soyal SM, Amleh A, Dean J. 2000. FIGα, a germ cell-specific transcription factor required for ovarian follicle formation. *Development* **127:** 4645–4654.
Stelkens RB, Wedekind C. 2010. Environmental sex reversal, Trojan sex genes, and sex ratio adjustment: Conditions and population consequences. *Mol Ecol* **19:** 627–646.
Sun M, Yang L, Feldman RI, Sun X, Bhalla KN, Jove R, Nicosia SV, Cheng JQ. 2003. Activation of phosphatidylinositol 3-kinase/Akt pathway by androgen through interaction of p85α, androgen receptor, and Src. *J Biol Chem* **278:** 42992–43000.
Teranishi M, Shimada Y, Hori T, Nakabayashi O, Kikuchi T, Macleod T, Pym R, Sheldon B, Solovei I, Macgregor H. 2001. Transcripts of the MHM region on the chicken Z chromosome accumulate as non-coding RNA in the nucleus of female cells adjacent to the DMRT1 locus. *Chromosome Res* **9:** 147–165.
Trapnell C, Pachter L, Salzberg SL. 2009. TopHat: Discovering splice junctions with RNA-Seq. *Bioinformatics* **25:** 1105–1111.
Vance SA. 1996. Morphological and behavioural sex reversal in mermithid-infected mayflies. *Proc R Soc Lond B Biol Sci* **263:** 907–912.
Wallace H, Badawy GM, Wallace BM. 1999. Amphibian sex determination and sex reversal. *Cell Mol Life Sci* **55:** 901–909.
Wang J, Saxe JP, Tanaka T, Chuma S, Lin H. 2009. Mili interacts with tudor domain-containing protein 1 in regulating spermatogenesis. *Curr Biol* **19:** 640–644.
Warner DA, Shine R. 2005. The adaptive significance of temperature-dependent sex determination: Experimental tests with a short-lived lizard. *Evolution* **59:** 2209–2221.
Warner DA, Shine R. 2008. The adaptive significance of temperature-dependent sex determination in a reptile. *Nature* **451:** 566–568.
Wu GC, Chiu PC, Lyu YS, Chang CF. 2010. The expression of *amh* and *amhr2* is associated with the development of gonadal tissue and sex change in the protandrous black porgy, *Acanthopagrus schlegeli*. *Biol Reprod* **83:** 443–453.
Xiang H, Zhu J, Chen Q, Dai F, Li X, Li M, Zhang H, Zhang G, Li D, Dong Y, et al. 2010. Single base-resolution methylome of the silkworm reveals a sparse epigenomic map. *Nat Biotechnol* **28:** 516–520.
Youngson NA, Whitelaw E. 2008. Transgenerational epigenetic effects. *Annu Rev Genomics Hum Genet* **9:** 233–257.
Zhuang Z, Wu D, Zhang S, Pang Q, Wang C, Wan R. 2006. G-banding patterns of the chromosomes of tonguefish *Cynoglossus semilaevis* Günther, 1873. *J Appl Ichthyology* **22:** 437–440.

Received June 17, 2013; accepted in revised form January 23, 2014.

Chapter Ⅳ High-resolution genetic map construction and its application in Japanese flounder

Shao CW, Niu YC, Rastas P, Liu Y, Xie ZY, Li HD, Wang L, Jiang Y, Tai SS, Tian YS, Sakamoto T, and Chen SL. Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. 2015, doi: 10.1093/dnares/dsv001.

DNA Research, 2015, 1–10
doi: 10.1093/dnares/dsv001
Original Article

Original Article

Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis

Changwei Shao[1,2,3,†], Yongchao Niu[4,†], Pasi Rastas[5], Yang Liu[1,2], Zhiyuan Xie[4], Hengde Li[6], Lei Wang[1,2], Yong Jiang[7], Shuaishuai Tai[4], Yongsheng Tian[1,2], Takashi Sakamoto[3,*], and Songlin Chen[1,2,*]

[1]Ministry of Agriculture, Yellow Sea Fisheries Research Institute, CAFS, Key Lab for Sustainable Development of Marine Fisheries, Qingdao 266071, China, [2]Function Laboratory for Marine Fisheries Science and Food Production Processes, National Lab for Ocean Science and Technology, Qingdao 266071, China, [3]Faculty of Marine Science, Tokyo University of Marine Science and Technology, Minato, Tokyo 108-8477, Japan, [4]BGI-Shenzhen, Shenzhen 518000, China, [5]Department of Biosciences, Metapopulation Research Group, University of Helsinki, Helsinki FI-00014, Finland, [6]Chinese Academy of Fisheries Science, Beijing 100039, China, and [7]National Oceanographic Center, Qingdao 266071, China

*To whom correspondence should be addressed. Tel. +86 0532-85844606. Fax. +86 0532-85811514. E-mail: chensl@ysfri.ac.cn; Tel. +81 03-5463-0450. Fax. +81 03-5463-0450. E-mail: takashis@kaiyodai.ac.jp

[†]These authors contributed equally.

Edited by Dr Toshihiko Shiroishi

Received 18 November 2014; Accepted 1 February 2015

Abstract

High-resolution genetic maps are essential for fine mapping of complex traits, genome assembly, and comparative genomic analysis. Single-nucleotide polymorphisms (SNPs) are the primary molecular markers used for genetic map construction. In this study, we identified 13,362 SNPs evenly distributed across the Japanese flounder (*Paralichthys olivaceus*) genome. Of these SNPs, 12,712 high-confidence SNPs were subjected to high-throughput genotyping and assigned to 24 consensus linkage groups (LGs). The total length of the genetic linkage map was 3,497.29 cM with an average distance of 0.47 cM between loci, thereby representing the densest genetic map currently reported for Japanese flounder. Nine positive quantitative trait loci (QTLs) forming two main clusters for *Vibrio anguillarum* disease resistance were detected. All QTLs could explain 5.1–8.38% of the total phenotypic variation. Synteny analysis of the QTL regions on the genome assembly revealed 12 immune-related genes, among them 4 genes strongly associated with *V. anguillarum* disease resistance. In addition, 246 genome assembly scaffolds with an average size of 21.79 Mb were anchored onto the LGs; these scaffolds, comprising 522.995 Mb, represented 95.78% of assembled genomic sequences. The mapped assembly scaffolds in Japanese flounder were used for genome synteny analyses against zebrafish (*Danio rerio*) and medaka (*Oryzias latipes*). Flounder and medaka were found to possess almost one-to-one synteny, whereas flounder and zebrafish exhibited a multi-syntenic correspondence. The newly

© The Author 2015. Published by Oxford University Press on behalf of Kazusa DNA Research Institute. This is an Open Access article distributed under the terms of the Creative Commons Attribution License (http://creativecommons.org/licenses/by/4.0/), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

developed high-resolution genetic map, which will facilitate QTL mapping, scaffold assembly, and genome synteny analysis of Japanese flounder, marks a milestone in the ongoing genome project for this species.

Key words: Japanese flounder, RAD-seq-based SNP, high-resolution linkage map, QTL mapping, genome synteny

1. Introduction

Japanese flounder (*Paralichthys olivaceus*), one of the most desirable and highly priced marine fish species, is widely cultured along the coast of Northeast Asian countries such as China, Japan, and Korea.[1] In addition to a booming Chinese market, >60,000 tons of Japanese flounder were artificially produced in Japan and Korea in 2013,[2] making *P. olivaceus* one of the most important species in stock enhancement programs in these countries. As a result of long-term, high-intensity stocking and resource management, however, farmed Japanese flounder have depressed immune systems that enhance individual susceptibility to microbial infections.[3] To increase profitability and sustainability while maintaining genetic variability in the cultured stock, the development of genetic breeding programs, such as marker- and gene-assisted selection, is urgently required. These approaches involve selection of genomic loci or genes related to the economic trait of interest using genetic maps constructed by molecular markers.[4]

An accurate genetic linkage map is therefore foundational to genetic breeding of a species.[5] To date, linkage maps have been constructed for at least 28 fish species; in addition, economically important traits have been mapped using these maps in over 13 fish species, including tilapia, rainbow trout, channel catfish, Atlantic salmon, common carp, and Asian seabass.[6] Following the construction of the first generation of Japanese flounder genetic maps based on simple sequence repeat (SSR) and amplified fragment length polymorphism (AFLP) markers by Coimbra *et al.*[7] and SSRs alone by Kang *et al.*,[8] the second generation of linkage maps has been developed with various types of molecular markers, such as SSRs and a few single-nucleotide polymorphism (SNP) markers, either alone or in combination.[9,10] Such efforts have resulted in the identification of several quantitative trait loci (QTLs) for disease resistance and growth in Japanese flounder. For instance, a single major genetic locus associated with lymphocystis disease resistance has been detected[11] and its candidate gene *tlr2* identified,[12] making the development of a lymphocystis disease-resistant strain of Japanese flounder possible by marker-assisted selection.[13] In addition, four QTLs for growth rate are clustered on genetic linkage group (LG) 14, suggesting that these major loci contributing to growth trait variation are potentially useful for marker-assisted selection in future genetic breeding programs of Japanese flounder.[10] Although previous genetic maps of Japanese flounder have been successfully used to map QTLs for a few economic traits, the number of commonly available markers, restricted to hundreds to a few thousand, makes it difficult to carry out fine-scale mapping to shrink the genomic regions tightly associated with important traits. A genetic map of much higher resolution is thus urgently needed for Japanese flounder.

A high-resolution genetic linkage map is also an excellent tool for genome map construction, which in turn allows direct comparison of chromosomal organization and evolution.[14] For instance, *Caenorhabditis briggsae* chromosomes reconstructed by SNP-based genetic mapping were compared with the *Caenorhabditis elegans* genome. An almost complete conservation of synteny was observed, with many cases of perfect 1:1 orthologues found between these two distantly related species.[14] In addition, construction of a dense gene-based map of the butterfly species *Bicyclus anynana* enabled broad-coverage analysis of synteny with the lepidopteran reference genome; this analysis suggested strong conservation of gene assignments to chromosomes and numerous large- and small-scale chromosomal rearrangements.[15] The application of genetic mapping allows genome-level analysis to be extended to non-model species, conferring an advantage in the comparative analysis of chromosomal organization and evolution.

With the emergence of next-generation (massively parallel) sequencing and associated technological genotyping advancements such as reduced representation library sequencing techniques,[16] SNP markers, representing the most abundant source of variation in the genome, have become widely used for high-resolution genetic map construction; they are also applied for measurement of genetic diversity and investigation of population structure.[17] As a reliable, high-throughput, affordable method to reduce genomic complexity, restriction-site associated DNA tag sequencing (RAD-seq) has been particularly attractive for SNP discovery and genotyping.[18] RAD-seq technology has been successfully applied in the construction of genetic maps for various species, such as stickleback,[19] rainbow trout,[20] eggplant,[21] barley,[22] chickpea,[23] guppy,[24] and lupin.[25] Furthermore, the discovery that replicate parallel phenotypic evolution may be occurring in stickleback on a genome-wide scale[26] as well as the detection of population admixture and improved identification of hybrid and non-hybridized individuals using RAD-seq[20] suggest that this technique is applicable to population genomics and phylogeography.

Despite the availability of two generations of genetic maps, the large-scale discovery and utilization of SNPs, the most promising markers for providing sufficiently dense genetic maps,[23] has not been carried out in Japanese flounder. In the present study, we consequently performed a large-scale identification of genome-wide SNPs derived from RAD-seq of a Japanese flounder mapping population containing 2 parents and 216 offspring. We then used the results to construct a high-density (third generation) SNP-based map. We also successfully detected QTLs for *Vibrio anguillarum* disease resistance and their related genes in Japanese flounder. Finally, we used our SNP-based genetic map in conjunction with the *P. olivaceus* Genome Sequencing Project (PoGSP) database, which is to be published in the near future. In particular, the map was used to facilitate the anchoring and orienting of scaffolds generated by whole genome sequence data assembly and for chromosomal-level comparative analysis.

2. Materials and methods

2.1. Mapping population

Twenty-four full-sib families of Japanese flounder were bred by Haiyang Yellow Sea Fisheries Co. (Yantai, China) in 2010.[27] All 216 offspring of one of the families, with an average length of 12.61 ± 1.40 cm and average weight of 18.85 ± 6.28 g, were challenged through intraperitoneal injection with a 0.2-ml bacterial suspension of $\sim 7.16 \times 10^5$ colony-forming units (CFU) of *V. anguillarum*.

Another 20 individuals were injected with 1× phosphate-buffered saline as a control. The median lethal concentration had been determined in our laboratory previously.[28] The offspring were kept in three 0.28-m^3 tanks supplied with fresh seawater at 23 ± 0.5°C. This challenge experiment was performed once and lasted for ~1 week. Mortality was recorded every 3 h based on the appearance of dead fish (Supplementary Table S1). Genomic DNA was isolated from the fins of the parental fish and 216 offspring using traditional phenol–chloroform extraction in combination with RNase treatment and stored at −20°C.[10] Before construction of RAD-seq libraries, all DNA samples were quantified using a NanoDrop instrument (Thermo Scientific, Wilmington, DE, USA), and their concentrations were adjusted to 50 ng/μl using Tris–EDTA buffer.

2.2. RAD library construction and sequencing

RAD-seq libraries were constructed using a protocol adapted from Baird *et al.*[19] Briefly, 250 ng of genomic DNA from each of 218 library individuals was digested separately with 20 units of PstI and then heat inactivated at 65°C. Various P1 adapters, each with a unique 4–8 bp molecular-identifying sequence (MID), were then ligated to designated individuals, which were then pooled in groups of 24 individuals and randomly sheared to DNA fragments. Sheared DNA was purified, eluted, and separated using gel electrophoresis, and a DNA fraction corresponding to 300–700 bp was excised and purified. After end repair, purification, and elution, dATP overhangs were added to the DNA fraction. A paired-end P2 adapter containing T overhangs was ligated to 20 μl of sheared, size-selected, P1-ligated, and pooled DNA template with a specific adapter. The ligated material was then purified, eluted, and subjected to PCR enrichment. Sequencing of the RAD products from the 218 individuals was performed on a HiSeq2000 next-generation sequencing platform. Sequencing data for each individual were then extracted according to the specific MID.

2.3. SNP discovery and genotyping

We first filtered out Illumina short reads lacking sample-specific MIDs and expected restriction enzyme motifs. All the short reads from each of the samples were then clustered into tag reads on the basis of sequence similarity (allowing one mismatch at most between any two reads within each tag read cluster, with clusters having <2 or >100 reads discarded).[29] Tag reads from the two parents were compared and filtered to remove monomorphic DNA sequences, leaving only the tag reads with SNPs. SNPs were selected according to the following criteria: (i) alleles with a minimum coverage of five reads and a score >20 ($P > 0.05$) were selected; (ii) the base at the SNP site was unique (if at a homozygous locus) or bimorphic (if at a heterozygous locus, with the minor base represented by at least three reads); and (iii) the ratio of the two kinds of bases (major to minor) at a heterozygous locus ranged from 1 to 5. Regions containing these putative SNPs were used as reference SNP regions. The clean tag reads of offspring individuals were then aligned to reference SNP regions, with genotypes of individuals determined by reference to the parental genotypes. Filtering of reads and SNP selection were performed following previously described algorithms[29,30] as implemented using custom scripts (available from the authors upon request).

2.4. Genetic map construction

A pseudo-testcross population was used to construct the linkage map. For the linkage analysis, RAD-based SNPs were first tested against the expected segregation ratio. Two SNP alleles heterozygous with respect to the two parents were expected to segregate in a 1:2:1 ratio, whereas a pair of SNP alleles, one heterozygous and the other homozygous, were expected to segregate in a 1:1 ratio. To construct a genetic map, markers showing significant segregation distortion ($P < 0.01$, χ^2 test) were removed. The remaining SNPs were then used to construct the genetic map with the recently developed software package Lep-MAP.[31] The genotype data were first filtered manually to remove obvious Mendelian errors from the offspring. LG assignments were then obtained using the separate chromosomes module with a logarithm of odds (LOD) score limit of 10. The marker order was obtained using the order markers module of Lep-MAP. Because this module uses only paternally informative markers in the ordering, a second genotype data file was constructed by swapping the parents of the original data. An integrated map was constructed by using both genotype files as input, with paternal and maternal maps generated by using only one input file. To speed up the computation, constant rates for genotype errors and recombinations were used. Finally, the marker positions and error parameters were established by order markers module.

2.5. QTL mapping for *V. anguillarum* disease resistance

QTL analyses were conducted with the WinQTLCart2.5 software program using the composite interval mapping (CIM) method.[32] The CIM analysis was run using Model 6 with four parameters for forward and backward stepwise regression, a 10-cM window size, five control markers, and a 1-cM step size. The LOD threshold value was determined on the basis of 1,000 permutations at a whole genome-wide significance level of $P < 0.05$. The location of each QTL was determined according to its LOD peak location and surrounding region. Other QTL reference values, including phenotypic variation and a positive or negative additive effect for *V. anguillarum* disease resistance, were also calculated by WinQTLCart2.5. Candidate genes for *V. anguillarum* disease resistance were identified by mapping the corresponding tags of SNPs in QTL regions to the scaffold assembly followed by retrieval of the corresponding gene ID from the gene annotation file.

2.6. Genome scaffold assembly and genome synteny

We used the high-resolution RAD-based SNP genetic map and BLASTN (E-value < 1×10^{-5}, identity ≥95%, and alignment rate >50%) to map reads containing SNPs with less than two mismatches to PoGSP-derived scaffolds. To reduce the complexity of scaffolds anchored to hundreds of SNP markers, a tag SNP was selected from each scaffold with multiple SNPs. The position and orientation of scaffolds were ordered and assembled onto 24 pseudo-chromosomes corresponding to the 24 LGs based on genetic distances between SNPs. Scaffolds with only one SNP marker could be anchored but not oriented because of the lack of markers. We then used the Japanese flounder reference gene set generated from the PoGSP database along with medaka (*Oryzias latipes*) and zebrafish (*Danio rerio*) gene sequences from Ensembl (release 57). After filtering short genes (coding sequence <150 bp), we chose the transcripts with the longest coding sequences to represent each gene and then identified reciprocal best-matching orthologous genes between Japanese flounder and the other fish using BLASTP (E-value = 1×10^{-10}). Syntenic regions defined by the top hits of the homology search were plotted using an in-house script.

3. Results

3.1. RAD-seq library construction and sequencing

A total of 218 RAD-seq libraries from 2 parents and their 216 offspring were constructed and sequenced on an Illumina HiSeq2000 platform to generate 3.91 billion raw reads. After data trimming, 3.52 billion reads, comprising ~149.87 Gb of sequencing data, were individually partitioned into RAD tags according to their MIDs. Finally, female and male parental data sets, containing respectively 22.93 million filtered reads (comprising 947.96 Mb of data with a GC% of 47.02) and 17.7 million filtered reads (comprising 796.52 Mb of data with a GC% of 47.21), were correspondingly partitioned into 6,294,857 and 4,843,594 RAD tags. From the 216 offspring, a total of 3.48 billion filtered reads (average of 16.12 million) corresponding to 148,128.63 Mb of data (average of 685.78 Mb) were produced and divided into 939,770,718 RAD tags (ranging from 1,627,621 to 8,329,563 with an average of 4,350,790) for individual SNP discovery (Supplementary Table S1).

3.2. SNP discovery and genotyping

After stringent selection according to the above-described method, RAD tags from each individual were clustered and compared. RAD tags containing SNPs of the two parents were identified using a custom k-mer matching algorithm that excluded exact sequence matches (monomorphic loci) per 41-bp sequence. A panel of 13,362 high-fidelity SNPs with fixed genotypes in both parents was identified using these criteria described under 'Material and methods', and alleles for each marker were assigned to their respective parental donor. Using an in-house script, these SNPs were then analysed using genotyping data across the 216 offspring and classified into three categories: maternal heterozygous (6,741 SNPs), paternal heterozygous (4,128 SNPs), and heterozygous in both (2,493 SNPs). All 13,362 SNPs and their flanking sequences are listed in Supplementary Table S2.

3.3. High-resolution genetic map construction

A high-resolution RAD-based SNP genetic map of Japanese flounder based on a pseudo-testcross population was first constructed using Lep-MAP. A total of 12,712 segregating SNPs (95.14%) were successfully classified into 24 LGs (Table 1 and Fig. 1). The paternal map contained 8,784 SNPs with a total genetic distance of 2,561.17 cM; the length of each LG ranged from 19.53 cM (LG8) to 145.08 cM (LG19), with an average genetic length of 106.72 cM (Supplementary Fig. S1). The corresponding maternal map consisted of 6,343 SNPs representing a total length of 2,356.86 cM and ranging from 0.6 cM (LG8) to 187.33 cM (LG1) (Supplementary Fig. S2). More interestingly, LG8 was the smallest LG on both maternal (0.6 cM) and paternal maps (19.53 cM). LG8 on the maternal map contained only 3 SNPs heterozygous in both parents; on the paternal map, LG8 included 135 paternal heterozygous SNPs and 3 SNPs heterozygous in both parents. The resulting integrated map consisted of 24 LGs including 12,712 SNPs, which corresponded to 7,430 effective loci. The total map length was 3,497.29 cM, with an average inter-locus distance of 0.47 cM; the genetic length of each LG ranged from 21.64 cM (LG8) to 221.14 cM (LG20), with an average inter-locus distance of 0.37–0.75 cM. LG1 was the densest, having 495 effective loci with

Table 1. Characteristics of genetic maps and anchoring scaffolds of Japanese flounder

LG_ID	Paternal map		Maternal map		Integrated_Map					
	No. of SNPs	Distance (cM)	No. of SNPs	Distance (cM)	No. of SNPs	No. of effective loci	Distance (cM)	Average inter-loci distance	No. of anchored scaffolds	Length of anchored scaffolds (Mb)
1	493	137.15	561	187.33	875	495	208.74	0.42	16	27.810
2	75	46.95	94	9.03	127	72	36.33	0.50	9	18.596
3	451	111.49	375	137.75	706	407	190.32	0.47	10	23.206
4	321	97.84	284	95.24	527	290	142.28	0.49	7	18.984
5	413	101.92	216	80.38	584	303	114.34	0.38	10	25.846
6	441	126.54	442	181.24	742	440	203.21	0.46	9	24.619
7	224	94.54	76	72.64	253	178	94.76	0.53	13	29.554
8	138	19.53	3	0.60	138	59	21.64	0.37	8	11.100
9	377	103.89	46	9.00	407	195	101.09	0.52	19	14.865
10	222	88.72	151	76.56	333	214	123.10	0.58	12	21.850
11	263	82.11	385	127.91	573	351	177.62	0.51	8	22.769
12	452	102.60	461	121.51	787	436	161.24	0.37	10	25.650
13	410	130.09	412	107.02	686	401	172.96	0.43	5	19.776
14	313	117.59	260	88.67	421	276	128.28	0.46	11	11.147
15	486	122.12	151	91.38	509	297	150.71	0.51	10	23.798
16	175	72.03	294	100.37	402	253	119.09	0.47	12	17.162
17	408	136.93	492	169.11	717	462	209.14	0.45	11	25.188
18	276	116.96	99	84.30	287	192	144.18	0.75	8	19.412
19	458	145.08	174	97.10	484	302	169.02	0.56	12	26.392
20	528	114.33	464	159.35	824	467	221.14	0.47	6	22.917
21	449	127.47	103	32.21	492	247	101.46	0.41	6	25.309
22	448	102.11	260	106.97	590	380	157.75	0.42	10	19.992
23	514	141.20	220	101.53	603	336	173.14	0.52	12	24.253
24	449	121.98	320	119.66	645	377	175.75	0.47	12	22.800
Total	8,784	2,561.17	6,343	2,356.86	12,712	7,430	3,497.29	0.47	246	522.995
Average	366	106.72	264	98.20	530	310	145.72	0.47	10	21.79

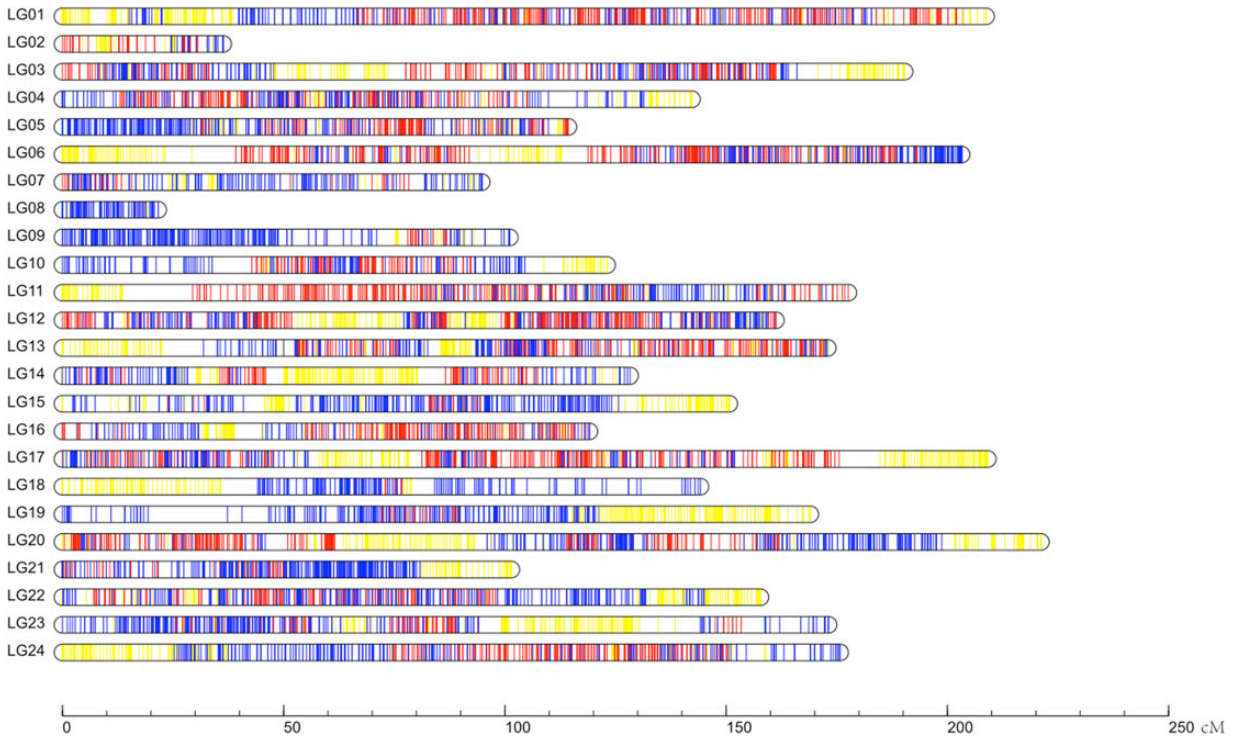

Figure 1. Linkage group lengths and marker distribution of the high-resolution restriction site-associated DNA sequencing-based SNP genetic map of Japanese flounder. Within each linkage group, red, blue, and yellow lines, respectively, represent maternal heterozygous SNPs, paternal heterozygous SNPs, and SNPs heterozygous in both parents. Genetic map details are given in Supplementary Table S3.

Table 2. Characteristics of *Vibrio anguillarum* disease resistance QTLs

QTL	Linkage group	Genetic position	Associated marker	LOD	[a]Exp%	Additive effect
qVA-1	LG6	89.8–90.8	record_231777.7	3	5.10	0.210012
qVA-2	LG6	95.9–99.3	record_245752.29	5	8.38	0.309202
qVA-3	LG6	100.7–102.5	record_247030.22	3.6	7.59	0.287123
qVA-4	LG6	104.1–105	record_246301.16	3.3	5.87	0.256509
qVA-5	LG6	105–105.4	record_254077.24	3.5	6.19	0.25732
qVA-6	LG6	107.8–108.7	record_254548.37	3.4	5.95	0.263431
qVA-7	LG19	115.6–115.8	record_255627.24	15.8	1.19	−0.116882
qVA-8	LG21	98.7–99.6	record_245734.7	3.7	6.06	−0.25436
qVA-9	LG21	100.6–100.9	record_255692.21	3.7	6.56	−0.273719
qVA-10	LG21	100.9–101.3	record_245651.14	3.3	6.36	−0.261304

[a]Exp, percentage of explained phenotypic variation.

an average density of 0.42 cM, whereas LG8 had the least number of effective loci (only 59). On average, each LG contained 310 effective loci spanning 145.72 cM (Table 1 and Fig. 1). Locus names and SNP positions on the 24 LGs of the integrated genetic map are listed in Supplementary Table S3.

3.4. *Vibrio anguillarum* disease resistance-associated QTLs and related genes

In total, 10 significant QTLs for *V. anguillarum* disease resistance were distributed on LG6, LG19, and LG21 of Japanese flounder (Table 2 and Fig. 2). Most of these QTLs were clustered together on their respective LGs. One major cluster containing six QTLs (qVA-1, qVA-2, qVA-3, qVA-4, qVA-5, and qVA-6) was detected between the narrow positions of 89.8–108.7 cM on LG6. Among them, qVA-2 located at 95.9–99.3 cM had the highest LOD value, 5, and correspondingly had the highest contribution to phenotypic variation, 8.38%. The other QTLs on LG6 were detected at positions 89.8–90.8, 100.7–102.5, 104.1–105.0, 105.0–105.4, and 107.8–108.7 cM, with LOD values of 3.0–3.6 and contributions to phenotypic variation of 5.1–7.59% (Supplementary Fig. S3). On LG21, another cluster situated within a short region (98.7–101.3 cM) consisted of three QTLs (qVA-8, qVA-9, and qVA-10) with a LOD value of 3.3–3.7 and was able to explain 6.06–6.56% of the phenotypic variation (Supplementary Fig. S3). Finally, qVA-7 containing a single SNP was centred around 115.6 cM on LG 19; it had a LOD value of 15.8 and explained 1.19% of the phenotypic variation. Although these 10 QTLs explained >59.25% of the total phenotypic variation, no major loci (explaining >20% of the total variation) were detected. The fact that these loci do not independently have higher contributions

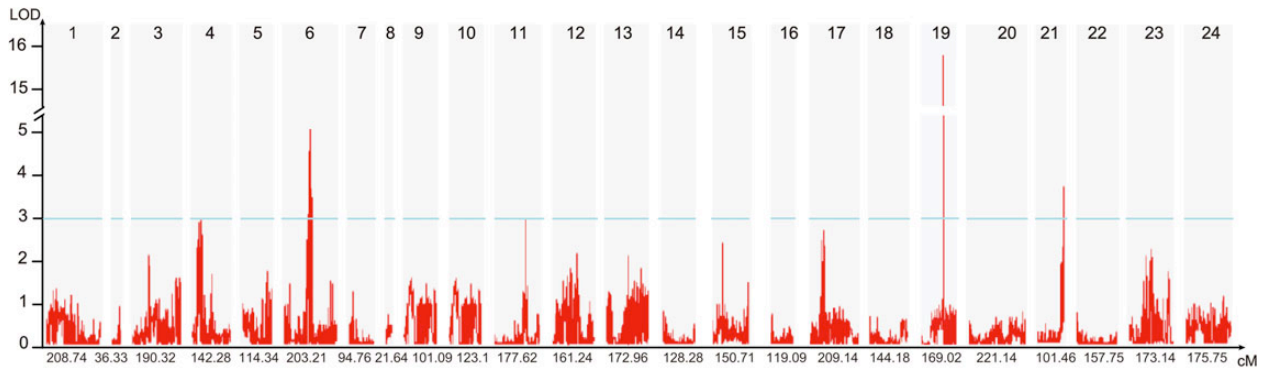

Figure 2. Genetic location of QTLs for *Vibrio anguillarum* disease resistance along the Japanese flounder genome. The blue horizontal line represents a linkage group-wise logarithm of odds (LOD) significance threshold of 3.0.

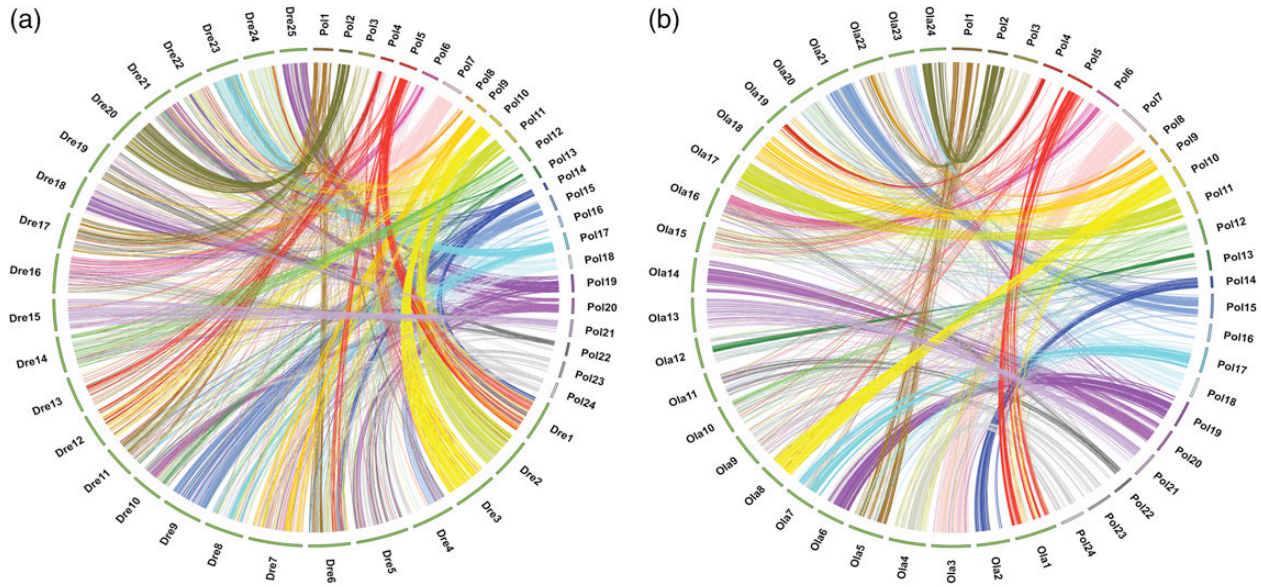

Figure 3. Circos diagram representing syntenic relationships between flounder and (a) zebrafish and (b) medaka, respectively. Each colored arc represents an orthologous match between two species. ~~Ola, *Oryzias latipes;* Dre, *Danio rerio;* Pol, *Paralichthys olivaceus*.~~

to such a complicated trait is not unexpected. Several genes related to *V. anguillarum* disease resistance were identified from the higher density map based on the Japanese flounder genome assembly. As shown in Supplementary Table S4, 218 genes were identified from the QTL regions, 12 of which were immune-related genes (*tap1*, *rftn1*, *satb1*, *cd40*, *cd69*, *aicda*, *mtss1*, *ccr4*, *azi2*, *mrc1*, *nod1*, and *tgfbr2*) functioning as key factors in different immune-gene signalling pathways or as antigen receptors. Interestingly, *tap1* and *satb1* act as molecular organizers for the major histocompatibility complex (MHC) class I,[33] while *cd40* and *cd69* are accessory molecules of MHC class II[34] directly involved in *V. anguillarum* disease resistance in Japanese flounder.

3.5. Genome assembly and synteny

The high-resolution genetic map was used in the present study for genome scaffold assembly. Among the 12,712 SNPs on the high-resolution map, 12,463 were successfully used to anchor 246 scaffolds representing 522.995 Mb to create a genome map of Japanese flounder. In total, we constructed 24 pseudo-chromosomes based on the LGs; each chromosome comprised an average of 21.79 Mb (10 scaffolds), corresponding to a mean linkage distance of 145.72 cM and suggesting a physical/genetic distance ratio of 149.53 Kb/cM. Among the LGs, LG7 holding 13 scaffolds corresponded to the largest chromosome (29.554 Mb) and represented 95.78% of the entire assembly, whereas the smallest LGs contained 8 scaffolds representing 67.85% of the assembly (16.36 Mb) (Table 1). In addition, 219 scaffolds covering 520.445 Mb were oriented by assigning multiple genetically separated markers located on each scaffold. The orientation of 27 scaffolds covering 2.55 Mb could not be determined, because only single markers were present on those scaffolds.

Following genome map construction, we compared the chromosomal orders of protein-coding genes of Japanese flounder on each scaffold with their counterparts in medaka and zebrafish. A total of 6,768 1:1 best orthologues between Japanese flounder and zebrafish were identified. As shown in Supplementary Table S5 and Fig. 3, 7 of the 24 chromosomes of flounder (Po. 4, 6, 10, 11, 12, 15, 18, 21, and 22) were in relatively conserved synteny with zebrafish chromosomes Dr. 12, 16, 3, 2, 14, 9, 24, 15, and 19 respectively. Many of the flounder chromosomes were found to have syntenic blocks with

more than one chromosome of zebrafish and vice versa. An example of the multi-syntenic correspondence between zebrafish and flounder genomes is that of Po. 20, which exhibited one of the highest degrees of synteny: Po. 20 had 519 hits on zebrafish chromosomes, most of which (171, 142, and 136) were located on Dr. 05, 10, and 21, respectively. Similarly, Dr. 06 was found to be highly syntenic with Po. 01, 05, and 24, whereas Po. 01, 05, and 24 were in synteny with Dr. 11, 01, and 22, respectively. With respect to flounder–medaka genome synteny, the alignment of 6,463 orthologues on the 24 flounder chromosomes revealed an obvious syntenic relationship with the 24 corresponding chromosomes of medaka. Except for some chromosomes with minor multi-chromosome hits (<50 genes), the results of this comparison suggest that flounder is more closely related to medaka than to zebrafish (Fig. 3).

4. Discussion

4.1. High-resolution genetic map construction using RAD-based SNPs

Genetic maps constructed using molecular markers are important for genomic and genetic analyses of individual species. SNPs are particularly attractive for genetic map construction, because they represent the most common type of DNA polymorphism in the genome and are amenable to high-throughput genotyping.[17] In the present study, ~950 million RAD tags were generated from 2 Japanese flounder parents and 216 offspring individuals. We discovered 13,362 novel SNPs evenly distributed throughout the entire Japanese flounder genome, providing thereby a large number of genetic variation resources for future genome selection and genome-wide association studies. Among the identified SNPs, 12,712 SNPs were successfully genotyped and assigned to 24 LGs corresponding to the 24 chromosome pairs of the Japanese flounder genome. The total length of the resulting map was 3,497.29 cM, with an average marker spacing of 0.47 cM. Four genetic maps of Japanese flounder have been previously constructed using various markers,[7–10] but those maps are characterized by relatively larger genetic intervals and/or non-uniform marker distributions. The most recent genetic map[10] containing 1,487 SSRs with an average interval of 1.22 cM was formerly the densest reported flatfish linkage map and worthy of acclaim. However, whole genome-level analyses such as genome-wide association studies were still not practical using this map because of the limits imposed by the types of markers and their associated genotyping methods.[23] The denser SNP-based genetic map constructed in this study thus has greatly expanded application. It represents a considerable improvement over previous flounder genetic linkage maps based on SSR and AFLP markers.

Most LGs on our genetic map were considered to be saturated with an even distribution of three types of SNPs, namely, maternal heterozygous, paternal heterozygous, and double heterozygous SNPs; a few LGs, such as LG8, LG9, LG11, LG18, LG19, were not, however, revealing a complex genome containing regions enriched in repeat sequences and transposable elements.[35] Interestingly, LG8 encompassed the smallest genetic distance, contained only paternal and double heterozygous markers, and corresponded to the smallest physical distance (11.1 Mb). The fact that LG8 on the maternal map had only markers heterozygous in both parents whereas the group on the paternal map had only paternal heterozygous and double heterozygous markers suggests specific suppression of homologous recombination in this LG. Similar phenomena have been documented in other species having recently evolved sex chromosomes with large pseudoautosomal regions.[36,37] To overcome this complication, additional markers must be developed and a larger mapping population is required to increase the density of the linkage map.

The Lep-MAP program used in this study offers several advantages over JoinMap,[38] currently the most popular linkage-mapping software package. Lep-MAP, unlike JoinMAP, is suitable for use with tens of thousands of markers. Furthermore, Lep-MAP is faster on large data sets, requires no manual input, and handles genotyping errors better because it estimates genotyping error rates for each marker. Since its development in 2013, Lep-MAP has thus been heavily used for genetic map construction, such as for the high-resolution genetic map of *Glanville fritillary*.[39] Finally, the RAD-seq technique is proving to be a highly valuable tool for SNP discovery and genotyping followed by genetic map construction and has currently been applied to various species. The success of RAD-seq is owing to advances in next-generation technologies, which reduce the cost of DNA sequencing. Genotyping by sequencing is therefore expected to become increasingly popular for high-throughput genotyping, genetic map construction, and genome analysis.[40]

4.2. *V. anguillarum* disease resistance-related QTLs and associated genes

The generated high-resolution genetic map allowed us to perform QTL fine mapping for an economically important trait of Japanese flounder. In the present study, 10 QTLs associated with *V. anguillarum* disease resistance were found to be distributed on three LGs (LG6, LG19, and LG21). Interestingly, most of the QTLs were concentrated within a narrow region (cluster) on the LGs. Six QTLs were clustered together (89.8–108.7 cM) on LG6 corresponding to assembly scaffolds, scaffold196 (1.078 Mb), scaffold103 (1.062 Mb), scaffold31 (4.399 Mb), and scaffold36 (11.373 Mb), and three were found in another cluster (98.7–101.3 cM) on LG21, corresponding to scaffold190 (0.973 Mb) and scaffold70 (2.590 Mb). The quite small genetic and physical distances among QTLs in certain clusters suggest that the individual clusters are highly effective QTLs. One exception is qVA-7 on LG19. Even though it had the largest LOD value (15.8), qVA-7 contained only one SNP located in a physical gap between two scaffolds, indicating that it may be a false-positive QTL or, alternatively, the result of an error arising during genetic map construction due to segregation distortion. In addition, the smallest significant QTL (other than qVA-7) explained 5.1% of the phenotypic variation, whereas the maximum contribution was only 8.38%, reflecting the complexity of this polygenically controlled disease trait in Japanese flounder. Although the possibility exists that additional disease-controlling QTLs are present, the positive QTLs detected in this study, taken together, explained 58.06% of the total disease resistance variation, suggesting that these two QTL clusters play major roles. On the other hand, the results of our QTL analysis are not consistent with a previous report that the major QTL for *V. anguillarum* disease resistance is located on LG11 (LG7 in that study) in Japanese flounder.[41] In the previous study, two *V. anguillarum* disease resistance-associated polymorphic SSR markers were selected using bulked segregant analysis and then directly anchored onto LG11. The QTLs were finally identified by genotyping of 22 SSRs of LG11 in an F_1 population.[41] Because the QTL detection analysis was restricted to a specific LG at the very start, other potential QTLs may have been missed. The other possible explanation for the observed discrepancy is that it was the inevitable result of using a different family under different environmental conditions, just as repeated analysis of the same family treated by two successive challenges would result in lower repeatability for

QTL detection for such a complex quantitative trait.[42,43] To enhance the accuracy of QTL detection, results from multiple full-sib families should be integrated in the future.

To identify potential candidate genes, we compared the detected QTLs with the scaffold assembly and annotation of the Japanese flounder reference genome. A total of 12 immune-related genes were discovered, corresponding to the multi-population QTL region. The fact that four of these genes (*tap1*, *satb1*, *cd40*, and *cd69*) are associated with *V. anguillarum* disease resistance in Japanese flounder is unsurprising, as they are involved in the molecular function of the MHC class-dependent pathway of antigen presentation. Extensive studies have shown that MHC class I and II molecules play a pivotal role in immune defence systems, because they allow T cells to distinguish self from non-self.[44] In particular, investigations of allelic polymorphism and the pattern of evolution in MHC genes in Japanese flounder have indicated an association between MHC genes and *V. anguillarum* disease resistance.[28,45–47] These latter results also support the quality of our RAD data, because a key issue for QTL detection and associated gene analysis, given adequate genome coverage, is marker quality.

4.3. Comparative genome analysis

Next-generation sequencing technology has enabled generation of draft genomes for individual species as well as rapid development of comparative analyses among multiple genomes based on chromosomal-assembly levels.[48] Nevertheless, the large amount of repetitive DNA sequences and high heterozygosity levels in most species, especially marine species, has hindered assembly accuracy and also prevented direct assembly into chromosomal performance.[49] Consequently, a high-quality genetic map is an appropriate way to anchor scaffolds onto chromosomes. In this study, 24 pseudo-chromosomes of Japanese flounder were constructed by anchoring 246 scaffolds totaling 522.995 Mb, accounting for 95.78% of assembly sequences, using a high-resolution SNP genetic map. Such a high level of chromosomal-level genome map integrity is due to the SNPs, which were abundantly distributed across the genome by using the highest-throughput genotyping technique available.[17] The constructed pseudo-chromosomes of Japanese flounder provide a basic foundation to carry out extensive comparative genomics. In the present study, we performed an *in silico* analysis of orthologous gene pairs, because genes are evolutionarily conserved relative to intergenic regions.[50] From a general point of view, the orthologues (6,768) of flounder-medaka were more abundant than the orthologues (6,463) of flounder–zebrafish, reflecting the relatively closer phylogenetic relationship of medaka and flounder.[51] Looking more closely at evolutionary relationships based on the chromosomal segments, multi-syntenic correspondences were exhibited between flounder and zebrafish, whereas almost all flounder and medaka chromosomes had a 1:1 correspondence. This difference in synteny directly reflects that the flounder have a phylogenetic closeness to medaka than to zebrafish and also suggests that a much larger number of evolutionarily significant events, such as fusion, breakage, and translocation, have occurred in the zebrafish lineage. The medaka–flounder comparative analysis uncovered little evidence of such inter-chromosomal rearrangements in the medaka lineage despite the difference in chromosome numbers among those fish species (zebrafish has 25 chromosomes, whereas flounder and medaka have 24).[52–54] Furthermore, we inferred the occurrence of many intra-chromosomal events that have interrupted syntenic linkage blocks both between flounder and medaka and between flounder and zebrafish; these interruptions may have been the result of gene loss or divergence after duplication, which can contribute to deviations from synteny.[55] Most teleost species have experienced three rounds of whole-genome duplication. Thousands of duplicated genes remaining after divergence of teleost species have subsequently undergone non-reciprocal loss, pseudogenization, or sub-functionalization, and may be associated with unequal microsynteny in flounder and zebrafish.[56,57] From a different perspective, most chromosomal regions among the three analysed teleost species generally had a conserved syntenic relationship. This observation not only supports the marker ordering on the map but also facilitates the functional inference of genes in flounder and further benefits the analysis of genome evolution and comparative genomics.

In conclusion, we used RAD-seq technology for large-scale identification of SNPs that were then successfully used for high-throughput genotyping and construction of a high-resolution genetic map of Japanese flounder. The developed genetic map is the most comprehensive genetic map to date for this species. Through SNP mapping analysis, we identified 9 positive QTLs for *V. anguillarum* disease resistance that will be of interest to achieve breeding goals for Japanese flounder. We also anchored the genome scaffolds to pseudo-chromosomes and further identified complex syntenic relationships among flounder, medaka, and zebrafish by comparative genomic analysis. The large numbers of generated SNPs and the dense genetic map, coupled with future re-sequencing of multiple breeding families, should not only lay a foundation for chromosomal-level analysis of the flounder genome but should also provide an excellent resource for future molecular breeding efforts such as genome selection.

Supplementary data

Supplementary data are available at http://www.dnaresearch.oxfordjournals.org.

Funding

The work was supported by grants from State 863 High-Technology R&D Project (2012AA10A408), National Nature Science Foundation of China (31461163005), JSPS RONPAKU (Dissertation PhD) Program for C.S., and Taishan Scholar Project Fund of Shandong of China.

References

1. Fujiwara, A., Fujiwara, M., Nishida-Umehara, C., Abe, S. and Masaoka, T. 2007, Characterization of Japanese flounder karyotype by chromosome bandings and fluorescence in situ hybridization with DNA markers, *Genetica*, **131**, 267–74.
2. Nystoyl, R. and Tveteras, R. 2013, Fish production estimates and trends 2012–2013. Available at: http://www.gaalliance.org/cmsAdmin/uploads/tveteras.pdf.
3. Seikai, T. 2002, Flounder culture and its challenges in Asia, *Rev. Fish. Sci.*, **3–4**, 421–32.
4. Chapman, M.A., Pashley, C.H., Wenzler, J., et al. 2008, A genomic scan for selection reveals candidates for genes involved in the evolution of cultivated sunflower (*Helianthus annuus*), *Plant Cell*, **20**, 2931–45.
5. Andriantahina, F., Liu, X. and Huang, H. 2013, Genetic map construction and quantitative trait locus (QTL) detection of growth-related traits in *Litopenaeus vannamei* for selective breeding applications, *PloS ONE*, **8**, e75206.
6. Yue, G.H. 2013, Recent advances of genome mapping and marker-assisted selection in aquaculture, *Fish Fish.*, **15**, 376–96.

7. Coimbra, M.R., Kobayashi, K., Koretsugu, S., et al. 2003, A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*, *Aquaculture*, **1–4**, 203–18.
8. Kang, J.H., Kim, W.J. and Lee, W.J. 2008, Genetic linkage map of olive flounder, *Paralichthys olivaceus*, *Int. J. Biol. Sci.*, **4**, 143–9.
9. Castano-Sanchez, C., Fuji, K., Ozaki, A., et al. 2010, A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*), *BMC Genomics*, **11**, 554.
10. Song, W., Pang, R., Niu, Y., et al. 2012, Construction of high-density genetic linkage maps and mapping of growth-related quantitative trail loci in the Japanese flounder (*Paralichthys olivaceus*), *Plos ONE*, **7**, e50404.
11. Fuji, K., Kobayashi, K., Hasegawa, O., et al. 2006, Identification of a single major genetic locus controlling the resistance to lymphocystis disease in Japanese flounder (*Paralichthys olivaceus*), *Aquaculture*, **1–4**, 203–10.
12. Hwang, S.D., Fuji, K., Takano, T., et al. 2011, Linkage mapping of toll-like receptors (TLRs) in Japanese flounder, *Paralichthys olivaceus*, *Mar. Biotechnol.*, **13**, 1086–91.
13. Fuji, K.H.O., Honda, K., Kumasaka, K., Sakamoto, T. and Okamoto, N. 2007, Marker-assisted breeding of a lymphocystis disease-resistant Japanese flounder (*Pralichthys olivaceus*), *Aquaculture*, **1–4**, 291–5.
14. Hillier, L.W., Miller, R.D., Baird, S.E., et al. 2007, Comparison of *C. elegans* and *C. briggsae* genome sequences reveals extensive conservation of chromosome organization and synteny, *Plos Biol.*, **5**, e167.
15. Beldade, P., Saenko, S.V., Pul, N. and Long, A.D. 2009, A gene-based linkage map for *Bicyclus anynana* butterflies allows for a comprehensive analysis of synteny with the lepidopteran reference genome, *PLoS Genet.*, **5**, e1000366.
16. Moorthie, S., Mattocks, C.J. and Wright, C.F. 2011, Review of massively parallel DNA sequencing technologies, *HUGO J*, **5**, 1–12.
17. Berthier-Schaad, Y., Kao, W.H., Coresh, J., et al. 2007, Reliability of high-throughput genotyping of whole genome amplified DNA in SNP genotyping studies, *Electrophoresis*, **28**, 2812–7.
18. Rowe, H.C., Renaut, S. and Guggisberg, A. 2011, RAD in the realm of next-generation sequencing technologies, *Mol. Ecol.*, **20**, 3499–502.
19. Baird, N.A., Etter, P.D., Atwood, T.S., et al. 2008, Rapid SNP discovery and genetic mapping using sequenced RAD markers, *PloS ONE*, **3**, e3376.
20. Amish, S.J., Hohenlohe, P.A., Painter, S., et al. 2012, RAD sequencing yields a high success rate for westslope cutthroat and rainbow trout species-diagnostic SNP assays, *Mol. Ecol. Resour.*, **12**, 653–60.
21. Barchi, L., Lanteri, S., Portis, E., et al. 2012, A RAD tag derived marker based eggplant linkage map and the location of QTLs determining anthocyanin pigmentation, *PloS ONE*, **7**, e43740.
22. Chutimanitsakun, Y., Nipper, R.W., Cuesta-Marcos, A., et al. 2011, Construction and application for QTL analysis of a Restriction Site Associated DNA (RAD) linkage map in barley, *BMC Genomics*, **12**, 4.
23. Gaur, R., Azam, S., Jeena, G., et al. 2012, High-throughput SNP discovery and genotyping for constructing a saturated linkage map of chickpea (*Cicer arietinum L.*), *DNA Res.*, **19**, 357–73.
24. Willing, E.M., Hoffmann, M., Klein, J.D., Weigel, D. and Dreyer, C. 2011, Paired-end RAD-seq for de novo assembly and marker design without available reference, *Bioinformatics*, **27**, 2187–93.
25. Yang, H., Tao, Y., Zheng, Z., Li, C., Sweetingham, M.W. and Howieson, J. G. 2012, Application of next-generation sequencing for rapid marker development in molecular plant breeding: a case study on anthracnose disease resistance in *Lupinus angustifolius L*, *BMC Genomics*, **13**, 318.
26. Hohenlohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A. and Cresko, W.A. 2010, Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags, *PLoS Genet.*, **6**, e1000862.
27. Liu, F., Chen, S.L., Wang, L., et al. 2013, Analysis of growth performance and breeding value of 'Ping You No.1' Japanese flounder and selection of parents, *J. Fish. Sci. China*, **3**, 521–7.
28. Xu, T.J., Chen, S.L., Ji, X.S. and Tian, Y.S. 2008, MHC polymorphism and disease resistance to *Vibrio anguillarum* in 12 selective Japanese flounder (*Paralichthys olivaceus*) families, *Fish Shellfish Immun.*, **25**, 213–21.
29. Catchen, J.M., Amores, A., Hohenlohe, P., Cresko, W. and Postlethwait, J. H. 2011, Stacks: building and genotyping Loci de novo from short-read sequences, *G3-Genes Genom.Genet.*, **1**, 171–82.
30. Catchen, J., Hohenlohe, P.A., Bassham, S., Amores, A. and Cresko, W.A. 2013, Stacks: an analysis tool set for population genomics, *Mol. Ecol.*, **22**, 3124–40.
31. Rastas, P., Paulin, L., Hanski, I., Lehtonen, R. and Auvinen, P. 2013, Lep-MAP: fast and accurate linkage map construction for large SNP datasets, *Bioinformatics*, **29**, 3128–34.
32. Silva Lda, C., Wang, S. and Zeng, Z.B. 2012, Composite interval mapping and multiple interval mapping: procedures and guidelines for using Windows QTL Cartographer, *Methods Mol. Biol.*, **871**, 75–119.
33. Hinz, A., Jedamzick, J., Herbring, V., et al. 2014, Assembly and function of the MHC I peptide-loading complex are conserved across higher vertebrates, *J. Biol. Chem.*, **289**, 33109–17.
34. Cohen-Sfady, M., Nussbaum, G., Pevsner-Fischer, M., et al. 2005, Heat shock protein 60 activates B cells via the TLR4-MyD88 pathway, *J. Immunol.*, **175**, 3594–602.
35. Schaack, S., Choi, E., Lynch, M. and Pritham, E.J. 2010, DNA transposons and the role of recombination in mutation accumulation in *Daphnia pulex*, *Genome Biol.*, **11**, R46.
36. Bergero, R. and Charlesworth, D. 2009, The evolution of restricted recombination in sex chromosomes, *Trends Ecol. Evol.*, **24**, 94–102.
37. Natri, H.M., Shikano, T. and Merila, J. 2013, Progressive recombination suppression and differentiation in recently evolved neo-sex chromosomes, *Mol. Biol. Evol.*, **30**, 1131–44.
38. van Ooijen, J.W. 2006, Joinmap 4: software for the calculation of genetic linkage maps in experimental populations. Kyazma B.V., Wageningen, Netherlands.
39. Ahola, V., Lehtonen, R., Somervuo, P., et al. 2006, The *Glanville fritillary* genome retains an ancient karyotype and reveals selective chromosomal fusions in Lepidoptera, *Nature Commun.*, **5**, 4737.
40. Elshire, R.J., Glaubitz, J.C., Sun, Q., et al. 2011, A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species, *PloS ONE*, **6**, e19379.
41. Wang, L., Fan, C., Liu, Y., et al. 2014, A genome scan for quantitative trait loci associated with Vibrio anguillarum infection resistance in Japanese flounder (Paralichthys olivaceus) by bulked segregant analysis, *Mar. Biotechnol.*, **16**, 513–21.
42. Yu, Y.T., Zhang, J.M., Shi, Y.S., Song, Y.C., Wang, T.Y. and Li, Y. 2006, QTL analysis for plant height and leaf angle by using different populations of maize, *J. Maize Sci.*, **14**, 88–92 (in Chinese).
43. Le Bras, Y., Dechamp, N., Krieg, F., et al. 2011, Detection of QTL with effects on osmoregulation capacities in the rainbow trout (*Oncorhynchus mykiss*), *BMC Genet.*, **12**, 46.
44. Adams, E.J. and Luoma, A.M. 2013, The adaptable major histocompatibility complex (MHC) fold: structure and function of nonclassical and MHC class I-like molecules, *Annu. Rev. Immunol.*, **31**, 529–61.
45. Xu, T.J., Chen, S.L. and Zhang, Y.X. 2010, MHC class IIalpha gene polymorphism and its association with resistance/susceptibility to Vibrio anguillarum in Japanese flounder (*Paralichthys olivaceus*), *Dev. Comp. Immunol.*, **34**, 1042–50.
46. Xu, T.J., Sun, Y.N. and Chen, S.L. 2010, Allelic variation, balancing selection and positive selected sites detected from MHC class Ialpha gene of olive flounder, *Genetica*, **138**, 1251–9.
47. Du, M., Chen, S.L., Liang, Y., et al. 2011, Polymorphism and balancing selection of MHC class II DAB gene in 7 selective flounder (*Paralichthys olivaceus*) families, *Evid. Based Complement. Alternat. Med.*, **20,11**, 613629.
48. Buermans, H.P. and den Dunnen, J.T. 2014, Next generation sequencing technology: advances and applications, *Biochim. Biophys. Acta.*, **10**, 1932–41.
49. Treangen, T.J. and Salzberg, S.L. 2012, Repetitive DNA and next-generation sequencing: computational challenges and solutions, *Nat. Rev. Genet.*, **13**, 36–46.
50. Jiang, Y., Gao, X., Liu, S., et al. 2013, Whole genome comparative analysis of channel catfish (*Ictalurus punctatus*) with four model fish species, *BMC Genomics*, **14**, 780.
51. Kluver, N., Pfennig, F., Pala, I., et al. 2007, Differential expression of anti-Mullerian hormone (amh) and anti-Mullerian hormone receptor type II (amhrII) in the teleost medaka, *Dev. Dynam.*, **236**, 271–81.

52. Dai, Z., Xiong, Y. and Dai, X. 2014, Neighboring genes show interchromosomal colocalization after their separation, *Mol. Biol. Evol.*, **31**, 1166–72.
53. Zhang, Y., Liu, S., Lu, J., et al. 2013, Comparative genomic analysis of catfish linkage group 8 reveals two homologous chromosomes in zebrafish and other teleosts with extensive inter-chromosomal rearrangements, *BMC Genomics*, **14**, 387.
54. Woods, I.G., Wilson, C., Friedlander, B., et al. 2005, The zebrafish gene map defines ancestral vertebrate chromosomes, *Genome Res.*, **15**, 1307–14.
55. Postlethwait, J.H., Woods, I.G., Ngo-Hazelett, P., et al. 2000, Zebrafish comparative genomics and the origins of vertebrate chromosomes, *Genome Res.*, **10**, 1890–902.
56. Glasauer, S.M. and Neuhauss, S.C. 2014, Whole-genome duplication in teleost fishes and its evolutionary consequences, *Mol. Genet. Genomics.*, **289**, 1045–60.
57. Brunet, F.G., Roest Crollius, H., Paris, M., et al. 2006, Gene loss and evolutionary rates following whole-genome duplication in teleost fishes, *Mol. Biol. Evol.*, **23**, 1808–16.

Conclusion

Flatfishes are good model teleost for the scientific research on sex determination and immune regulatory mechanism. In this thesis, we employed the half-smooth tongue sole and Japanese flounder, which exhibit a novel and intriguing mode of sex determination system and disease resistance, respectively, as target species. We firstly constructed two BAC libraries with a total of 55,296 BAC clones, corresponding to 13.36 times of genome in tongue sole. We then isolated the positive clones containing sex-related genes or female-specific markers, which further be performed to the sequencing analyzer. A BAC sequence assembly containing *cyp19a1a* gene were thoroughly analyzed, suggest a conserved structural characteristics and functional evolution. In order to understanding of the sex determination and differentiation in depth, we sequenced the male and female tongue sole separately and then analyzed the sex reversal mechanism using the whole-genome level of methylation analysis. Comparative analysis of the gonadal DNA methylomes of pseudomale, female, and normal male fish revealed that methylation modification in pseudomales is globally inherited in their ZW offspring, which can naturally develop into pseudomales without temperature incubation. The dosage compensation in a specific Z chromosomal region and special expression pattern of W chromosomal genes suggest that epigenetic regulation plays multiple crucial roles in sexual reversal of tongue sole fish. In Japanese flounder, we constructed a high-resolution genetic map using RAD-seq technology. 10 QTLs and their associated genes for *V. anguillarum* disease were found and the complex synenic relationships among three teleost (Japanese flounder, medaka and zebrafish) were detected at the whole genome level.

In conclusion, the development of BAC recourses and genome sequence as well as the methylation profiles in tongue sole provided a readily useable platform for genomics and epigenetics research. In particularly, this is the first comprehensive investigation of gene expression and epigenetic regulation at the whole genome level in a GSD species that exhibits the ESD phenomenon. A special phenomenon of transgenerational epigenetic inheritance of sexual reversal in tongue sole gives us an exciting example to understanding the epigenetic changes have been linked to sex determination in vertebrates, although we know that epigenetic marks can sometimes be transmitted to offspring in mammals. These novel insights on the roles of epigenetics from a distinctly non-model fish species would provide a reference for other non-model species. Besides, the large numbers of generated SNPs and the dense genetic map, coupled with future re-sequencing of multiple breeding families of Japanese flounder, should not only lay a foundation for chromosomal-level analysis of the flounder genome but should also provide an excellent resource for future molecular breeding efforts such as genome selection.

Acknowledgements

Firstly, I would greatly thankful to my supervisor Dr. Takashi Sakamoto for his nice discussion and suggestion on my research during these three years. Next I have to thank my lab director Dr. Songlin Chen for his support on my research. Also, I would like to thank my committee members Dr. Goro Yoshizaki and Dr. Ikuo Hirono for their helpful comments during my committee defense and Dr. Motohiko Sano for his constructive comments on my thesis. Special thank to my collaborator and friends Dr. Manfred Schartl, Dr. Francesc Piferrer, Dr. Pasi Rastas, Dr. Domitille Chalopin et al., your encouragement, and directions that making my study well and taking me to the right path. Especially to my wife, thanks for your support and understanding while I achieved another of my crazy endless personal and professional goals, and to my mother, thanks for always encourage me to follow my dreams, and to my daughter, you are my power. Finally, I would also like to thank the JSPS RONPAKU (Dissertation PhD) Program for providing me the scholarship to pursue my Doctor's degree.