Detection of QTL, genes and microRNA involved in sex determination and differentiation of Nile tilapia (*Oreochromis niloticus*)

Thesis submitted for the degree of

"Doctor of Philosophy"

By

Orly Eshel

Submitted to the Senate of the Hebrew University of Jerusalem

December, 2013

This work was carried out under the supervision of

Prof. Micha Ron and Prof. Gideon Hulata

Abstract

Fish of the genera *Oreochromis*, *Sarotherodona* and *Tilapia* of the family Cichlidae, known as tilapia, are among the most important cultured fish in tropical and subtropical countries. Tilapia production relies on monosex cultures of males, which so far proved difficult to maintain in large scale production facilities. Detection of sex determination (SD) genes in tilapia has both scientific and commercial importance.

The hypothesized sex chromosome system for *O. niloticus* is XX-XY based on breeding of hormonally sex-reversed individuals. For instance, crosses of sex reversed (ΔXX) males with normal (XX) females produced only females. Previous studies indicated that SD in tilapia is controlled by major genetic factors that may interact with minor genetic as well as environmental factors. Therefore SD should be analyzed as a quantitative trait. Quantitative trait loci (QTL) in tilapia affecting SD were previously mapped to linkage groups (LG) 1 and 3. Genetic markers associated with distorted sex ratios were detected on LG 2, 6 and 23. The QTLs had large confidence intervals, hence, containing dozens of genes. A previous study detected differentially expressed genes in XX and XY bi-potential gonads during the period of 9–10 days post fertilization (dpf). Additionally, microRNAs are highly conserved among eukaryotes and are considered important regulatory elements in many biological processes during development, such as cell growth, differentiation and apoptosis. Sexually dimorphic expression patterns of microRNA were found in several species.

The goal of this study was to fine map a previously detected SD QTL and to identify genes and microRNAs affecting SD in the early developing embryo of Nile tilapia. Thus, SD mechanism was explored in the developing embryo from 2 to 9 dpf to detect primary differential expression and regulation of tilapia SD.

A full-sib progeny of Nile tilapia (Swansea strain), was divided into three groups: (i) untreated, (ii) feminized by diethylstilbestrol and (iii) masculinized by 17a-methyltestosterone. The first group was analyzed for association of microsatellite markers representing these five LGs (1, 2, 3, 6 and 23). The strongest association with gender was found on LG23 for marker *UNH898* (X^2; $p=8.6x10^{-5}$). Allele 276 was found almost exclusively in males, and we hypothesized that this allele is a male-associated allele (MAA). Sex-reversed individuals were used for mating experiments with and without the segregating MAA. Mating of individuals lacking the MAA resulted in all-female progeny. Mating of two individuals, heterozygous for MAA resulted in 81 males and 30 females (ratio of 3:1 as expected from XX/XY SD system). Analysis of association between gender and genotypes identified the MAA in 98.6% of males as opposed to 8.0% of females (χ^2; $p=2.5x10^{-18}$).

Eight markers flanking *UNH898* were genotyped for fine mapping of the QTL on LG23. A 95% confidence interval spanning positions 16-21 cM was determined.

In a physical mapping study, a segregating family of 90 individuals was genotyped for additional microsatellite genetic markers. The sex associated region on LG23 was localized to scaffold 101 between markers *GM597* and *ARO124* from 990,577 to 2,468,000 bp. Twelve adjacent markers found in this region were homozygous in females and either homozygous for the alternative allele or heterozygous in males. Markers flanking the critical sex region were heterozygous in two females thus localizing the QTL on LG23 into a 1.47 Kbp region. This genomic region harbored 51 positional candidate genes including a single functional candidate gene *anti-Müllerian hormone* (*amh*). Narrowing down this interval, to a single cM unit, will require sampling of thousands of fish.

Tilapia genome sequencing data enabled us to search for SD genes and microRNA in a genome wide scale. The objective was to detect genes and regulatory elements involved in early SD development by differential expression between genders at 2, 5 and 9 dpf. Artificial fertilization of *O. niloticus* females with either sex-reversed males (ΔXX) or genetically-modified males ('YY') resulted in all-female and all-male embryos for the analysis of genes and microRNA expression. A total of 56 biological samples were subjected to a microarray experiment, and six pools of the same samples were used as template for small RNA sequencing. A custom Agilent eArray was designed on the basis of the tilapia genome-sequence gene predictions and EST libraries. The array consisted of 43,803 probes each of 60 bases length, and controls for genome-wide expression profiling. A total of 59 differentially expressed genes were significant at false discovery rate (FDR) of 5%. A total of 831 microRNA precursors were discovered, of which nine had sexually dimorphic expression patterns with deviations between genders $\geq$ 4 standard deviations (FDR of 1.7%).

The genes with the most significant overexpression in females were *carbonyl reductase-like 20β-hydroxysteroid dehydrogenase* (*cr/20β-hsd*; $p=1\times10^{-17}$), *reticulon-4-interacting protein 1 homolog* ($p=1\times10^{-17}$) and *inositol monophosphatase 1* ($p=1\times10^{-14}$). *Zinc-finger BED domain- containing 3* ($p=1\times10^{-12}$), *Tetraspanin-8* ($p=1\times10^{-16}$) and *Amh* ($p=1\times10^{-9}$) were the most significant overexpressed genes in males. The expression of *amh* and *cr/20β-hsd* was tested by qPCR in brain, liver and gonads of tilapia males and females at 75 dpf. *Cr/20β-hsd* and *amh* genes' overexpression were validated in the respective gonads, i.e. ovary and testis, while *amh* was also highly expressed in male brain.

Eight genes were sampled for validation by qPCR, and a correlation of 0.8 for expression ratio of genders was obtained between the microarray and qPCR. Nevertheless, melting curve analysis showed multiple amplified fragments that were specific to one of the genders for a few of the genes, indicating potential copy number variation. Further analysis by qPCR of genomic DNA of males and females for four of the eight genes: *amh*, *cr/20β-hsd, tetraspanin-8* and *glycoprotein-A33,*

demonstrated significant copy number differences between genders, supporting the qPCR results. All four genes showed positive correlation between number of copies and expression level.

amh was a positional and functional candidate gene from our previous findings, and highly expressed in males both in cDNA and DNA levels. Sequencing analysis of *amh* identified a male specific duplication of this gene, denoted *amhy*, differing from the sequence of *amh* by a 233 bp deletion on exon 7, hence lacking the capability to encode the protein motif that binds to the transforming growth factor beta receptor. This shorter fragment was also found in cDNA of testis but not of ovary.

Six of the nine differentially expressed microRNAs were male-specific previous to 9 dpf. The 3' UTR of the 59 genes differentially expressed between genders in the microarray experiment were tested as potential targets for nine microRNAs that were differentially expressed between genders. Only pma-mir-4585 that was up-regulated in males showed significant perfect inverse correlation of expression pattern with all six targeted genes that were down-regulated in males.

Two of the nine differentially expressed microRNAs were up-regulated in males and highly conserved among vertebrates. Therefore, their gene hosts could be determined by across-species genomic analysis. Mir-21 and mir-218 were identified in *Tubulin Delta 1* and *GTP binding protein 1*, respectively. Both genes are known to be expressed in testis and spermatogenesis which is consistent with the finding that microRNAs are usually coordinately co-expressed with their host genes mRNAs.

In summary, this study reports the fine-mapping of QTL on LG23, the discovery of sexually-dimorphic expression patterns of genes and microRNAs, and genes enriched for copy number variation and apoptosis. This is the first report of male-specific *amh* duplication and detection of SD pathways that are functional at 2 to 9 dpf tilapia embryos. Thus, the sequence of events leading to SD in tilapia is apparently initiated soon after fertilization. Genes involved in apoptosis were shown to be a substantial group among the differentially expressed genes and constitute targets for sexually-dimorphic microRNA. *Amh* and *cr/20β-hsd* genes and the apoptosis process have a major role in SD similar to the SD model proposed for zebrafish.

Table of contents

General introduction

Sex Determination Mechanisms

A sex-determination (SD) system controls the sexual characteristics of an organism. The most common SD phenotypes are the two differentiated sexes. There are also some hermaphrodite species or one sex due to parthenogenesis (female reproducing without fertilization). SD mechanism can be controlled by genetic factors, environmental factors or a combination of the two types. In animals sex is often determined by chromosomal differences. In many species, the presence of a dominant sex determiner controls the sex of the individual; a Y chromosome in an XX/XY system (mammals and fly) or a W chromosome in a ZZ/WZ system (birds and reptiles) (Ezaz *et al*., 2006; Martınez *et al*., 2009). Studies have shown interesting similarities between the two systems (XX/XY and ZZ/WZ) of sex-determination (Ellegren, 2011). The ratio of X chromosomes to autosomes (the X:A ratio) is another genetic factor that determines the sex of the individual (Deng *et al*., 2011) as in *Drosophila melanogaster* where sex is determined by the activation of the most upstream gene, Sex lethal (*Sxl*) in the SD pathway (Schutt and Nothiger, 2000; Pomiankowski *et al*., 2004).

In organisms such as crocodiles, turtles and some fish, sex is determined by environmental variables (such as temperature, pH and hormones) or social variables (the size of an organism relative to other members of its population) - reviewed in Volff and Schartl (2001) and Devlin and Nagahama (2002). Extreme culture conditions with high densities produce male-biased sex ratios (Gousset, 1990). For some species, such as alligators, with temperature dependent SD, exposure to temperatures lower than or equal to 30°C resulted in all-females and temperatures higher than or equal to 34°C yielded all-males (Ferguson *et al*., 1982). In European eels (*Anguilla anguilla*) growth rate coupled with density influence SD. Females grow slower as a result of allocating resources for the development of ovaries (Huertas and Cerda, 2006). Davey and Jellyman (2005) have speculated that sexual differentiation is strongly influenced by growth rate and proposed that individuals that are experiencing the same growth rates are of the same sex.

In mammals, the presence of the Y chromosome determines the sex of the individual. Genes on the Y chromosome contain information that is “sex specific” for males (Charlesworth *et al*., 2005). The Y chromosome contains the *sex-determining region Y* (*SRY*) gene encoding testis-determining factor. *SRY* is responsible for the male phenotype and spermatogenesis (Sinclair *et al*., 1990; Koopman *et al*., 1991). *SRY* is unique to mammals and seems to have arisen in that lineage about 150 MY ago. The sex chromosomes of birds, a WZ SD system, yield ZZ genotypes as males while WZ genotypes produce females. The chicken Z chromosome contains *doublesex* and *mab-3 related transcription factor 1* (*DMRT1*), a male SD gene, and this gene is absent on the W chromosome

(Graves and Shetty, 2001; Schartl, 2004). In other model organisms, genes closely related to *DMRT1*, such as *doublesex* in *Drosophila* and *mab-3* in *Caenorhabditis elegans*, are involved in male sexual development (Raymond *et al*., 1998). At a critical phase of development where sexual differentiation begins, expression of *DMRT1* is higher in males than in females (Raymond *et al*., 1999; Smith *et al*., 2003). Similar results have been observed in sex-reversed males, supporting the idea that *DMRT1* is necessary for gonadogenesis and is a possible candidate SD gene (Smith *et al*., 2003). These different sex chromosome systems (XX/XY and ZZ/WZ) have apparently evolved independently as different genes are responsible for SD in each system (Ayling and Griffin, 2003; Charlesworth *et al*., 2005; Charlesworth and Charlesworth, 2005). It was speculated that they may have arose from different autosomal chromosomes (Ayling and Griffin, 2003; Charlesworth *et al*., 2005). Interestingly, duck-billed platypus (*Orniithirhynchus anatinus*) carries multiple sex chromosomes that are observed in male karyotypes as X1X2X3X4X5Y1Y2Y3Y4Y5 (Rens *et al*., 2004). The X1 chromosome contains genes homologous to the mammalian (human) X chromosome, while the X5 contains genes homologous to the bird WZ pair (Ezaz *et al*., 2006). However, the *DMRT1* gene has been identified on X chromosome while the *SRY* gene has not been identified (Ezaz *et al*., 2006).

Sex Determination Mechanisms in Fish

There are more than 24,000 species of fish (Nelson, 1994). Research on fish SD has provided important insight into the plasticity of the sex-determination process in vertebrates since the biology and ecology of fish is particularly diverse and provides unique examples of sex-determination mechanisms, yet they possess many of the same processes and pathways that are used in other vertebrate systems. Fish also provide unique opportunities to investigate and test theoretical concepts of SD, ranging from evolutionary mechanisms to biochemical processes.

SD mechanisms described in fishes range from monogenic to polygenic, including systems with dominant sex-determining factors mixed with influences from autosomal chromosomes (Devlin and Nagahama, 2002; Schartl, 2004; Ezaz *et al*., 2006). Teleost species are an interesting model for SD research, with a variety of SD systems and capability of producing viable hybrids between closely related species having different SD systems (Mank *et al*., 2006). Additionally, Bellott (2010) compared zebrafish, tetraodon, pufferfish, and medaka genomes to mammalian X and avian Z chromosomes and reported that most orthologs to Z and X genes occupy separate portions of each fish genome. Identifying these regions and their harbored-genes function may shed light on SD mechanisms interaction and evolution. The following is a summary of knowledge for each species or group separately:

Zebrafish (*Danio rerio*)

Despite the broad use of zebrafish as a model for vertebrate development, its SD mechanism remains poorly understood. There has been controversy about whether the zebrafish has a male (XY) or female (WZ) heterogametic SD system (Uchida *et al*., 2002; Tong *et al*., 2010). However, the observations that sex ratio variation decreases substantially under selective pressure, and that it is mainly influenced by parental genotypes strongly indicate that sex is a genetic trait in zebrafish (Liew *et al*., 2012). Several molecular tools used for comparing the male and female genomes e.g., cytogenetic analyses on zebrafish karyotypes and comparative FluoMEP assays rejected the existence of differentiated sex chromosomes (Wallace and Wallace, 2003; Liew *et al*., 2012). Genome-wide linkage analysis, using more than 5,000 sequence-based polymorphic restriction site associated (RAD-tag) markers and population genomic analysis of more than 30,000 single nucleotide polymorphisms (SNPs) identified multiple sex-associated regions (Anderson *et al*., 2012). Integration of these results indicates that zebrafish SD is polygenic. Genes with strong influence on SD and/or gonad differentiation are distributed throughout the genome and the combination of their alleles determines the sex of the individual with multiple parallel sex determining pathways (Anderson *et al*., 2012; Liew *et al*., 2012).

One example of SD-associated pathway is apoptosis. Rodriguez-Mari *et al*. (2010) found that homozygous *fancl* mutants develop exclusively into males but the introduction of *p53* mutations rescue the sex-reversal phenotype, allowing *fancl* mutants to become fertile females. The *Fancl* gene mediates cellular responses to a variety of stresses, including signals of DNA damage and apoptosis. It is also involved in the survival of developing oocytes through meiosis. Sex reversal of *fancl* mutants is not due to the absence of germ cells, but to an abnormal increase of germ cell apoptosis that compromises the survival of developing oocytes, and masculinizes the gonads.

Medaka (*Oryzias* spp.)

Over 20 different species are recognized in the genus *Oryzias* that is widely used as a model organism in developmental biology. Medaka species consist of XX/XY or ZZ/WZ SD system that can be modified by high temperatures, which cause XX females to develop into sex-reversed males (Matsuda, 2005; Sato *et al*., 2005). Various SD systems have been identified in seven species with linkage group (LG) corresponding to the sex chromosomes: *O. latipes* (LG1), *O. curvinotus* (LG1), *O. luzonensis* (LG12), *O. minutillus* (LG8), *O. dancena* (LG10), *O. hubbsi* (LG5), and *O. javanicus* (LG16) (Matsuda, 2005; Takehana *et al*., 2007, 2008; Nagai *et al*., 2008). Most medaka species have XX/XY SD systems whereas *O. hubbsi* has ZZ/ZW system. Like in many other fish species, no

cytogenetically distinct sex chromosomes have been observed in medaka (Schartl, 2004; Matsuda, 2005).

A DM-domain gene (*dmy*) was the first SD gene identified in a non-mammalian vertebrate, in *O. latipes* (Matsuda *et al*., 2002, 2007). During a critical developmental phase, the *DMY* gene is required for sex differentiation (Schartl, 2004). *Dmy* was found additionally in *O. curvinotu* but has not been detected in any other type of fish, including other *Oryzias* fishes (Kondo *et al*., 2003; Matsuda, 2005). Analysis of the Y-specific region of the *O. latipes* sex chromosome has demonstrated that *DMY* arose from duplication of the autosomal *dmrt1* gene (Nanda *et al*., 2002; Kondo *et al*., 2006). Myosho *et al*. (2012) found a Y-chromosome specific copy of gonadal soma-derived factor gene (*gsdf*), denoted *gsdfy*, in *O. luzonensis*. The *gsdf* gene is a member of the transforming growth factor β family (TGF-β). Overexpression experiments using a *gsdfy* genomic clone injected into one-cell–stage embryos of *O. luzonensis* resulted in all-male progeny, whereas all fish without the transgene developed as females.

Salmonids

The salmonid group consists of more than 66 species with 92% similarity at the DNA level (Nelson, 2006; Koop *et al*., 2008). Crosses between hormonally treated individuals and their progeny testing demonstrated that several salmonid species have XX/XY SD system (Devlin and Nagahama, 2002). However, sex ratio can be altered by additional sex modifying loci (autosomal regions) and environmental factors such as temperature (Azuma *et al*., 2004). The sharing of sex-linked markers on the X and Y chromosomes and the difficulty in identifying Y-specific markers indicate that X and Y chromosomes in salmonids have a large pseudoautosomal region and a small sex determining region. Linkage analyses suggest that either the SD region differs in different lineages or has remained intact and moved by transposition to different chromosomes (Davidson *et al*., 2009). Yano *et al*. (2012) have recently identified a sexually dimorphic transcript, named *sdY* on the Y chromosome-specific genomic sequence, OmyY1 in rainbow trout (*Oncorhynchus mykiss*) that displays sequence homology with the carboxy-terminal domain of interferon regulatory factor 9 (*Irf9*). *Irtf9* differs from *sdY* mainly by DNA-binding domain deletion and it was shown that the *sdY* gene is both necessary and sufficient to trigger testicular differentiation.

Fugu (*Takifugu* spp.)

Fugu is a large marine teleost with XX/XY sex determining system (Kikuchi *et al*., 2007). Genome-wide linkage mapping has shown that the SD region is restricted to a small segment of chromosome 19 flanked by large autosome-like regions. Kamiya *et al*. (2012) fine mapped this

region and found that a SNP causing change of an amino acid (His/Asp384) in the kinase domain of anti-Müllerian hormone receptor type II (*amhr2*) was fully associated with phenotypic sex in *Takifugu rubripes*, *T. pardalis* and *T. poecilonotus*.

Stickleback (*Gasterosteus* spp.)

Stickleback, one of five genera of the family Gasterosteidae, is a small, elongated fish found in fresh, brackish, and marine waters. It is a model organism with highly ritualized reproductive behavior. One of stickleback (Gasterosteidae) research aspect is the evolution of SD and sex chromosomes (Ross *et al*., 2009).

The first cytogenetic survey in this family reported heteromorphic XY pair in the black-spotted stickleback (*G. wheatlandi*) and a heteromorphic ZW pair in the four-spine stickleback (*Apeltes quadracus*) (Chen and Reisman, 1970). Ross *et al*. (2009) used genetic mapping and molecular cytogenetics to characterize the sex-chromosome systems of multiple stickleback species (Gasterosteidae) that confirmed earlier findings. They reported that male three-spine stickleback (*G. aculeatus*) have a heteromorphic XY pair corresponding to LG19. The nine-spine stickleback (*Pungitius pungitius*) has a heteromorphic XY pair corresponding to LG12. In black-spotted stickleback (*G. wheatlandi*) males, one copy of LG12 has fused to the LG19-derived Y chromosome, giving rise to an X1-X2-Y sex-determination system. The sex-chromosome diversity that was uncovered in sticklebacks provides a rich comparative resource for understanding the mechanisms that underlie the rapid turnover of sex-chromosome systems. Kitano *et al*. (2009) showed that a newly-evolved sex chromosome contains genes that contribute to speciation in three-spine stickleback (*G. aculeatus*). A neo-X chromosome contains loci for male courtship displaying traits that contribute to behavioural isolation (originated from LG19), whereas the ancestral X chromosome contains loci for both behavioral isolation and hybrid male sterility (originated from LG9). These results strongly suggest the contribution of sex-associated regions to reproductive isolation between closely related stickleback species.

Patagonian pejerrey (*Odontesthes hatcheri*)

Odontesthes hatcheri is a South American gonochoristic fish with both temperature-dependent and XX/XY SD system (Koshimizu *et al*., 2010). Hattori *et al*. (2012) reported that males carry a duplicated copy of the *anti-Müllerian hormone* homolog (*amh*) gene with 577 bp insertion in intron 3. The injection of an *amh* antisense morpholino (MO) in XY embryos resulted in development of ovaries in 11 of 50 larvae. Analyses performed during early stages of embryonic and larval

development revealed a comparatively early mRNA expression in relation to other teleosts (Fernandino *et al*., 2008).

Platyfish (*Xiphophorus maculatus*)

An unusual sex chromosome combination arises in the platyfish (*Xiphophorus maculatus*), which carries both X, Y, and W chromosomes (Schartl, 2004). In this species WY, WX, and XX genotypes yield females, while XY and YY genotypes develop into males (Schartl, 2004). The presence of the female-determining W chromosome, when paired with either the Z or Y chromosome, always determines the sex of the individual into female (Devlin and Nagahama, 2002).

Evolution and CNV Aspects of Sex Chromosome Systems

Chromosomal sex determination is common among animals. In most vertebrate species with genetically determined sex, no differentiated sex chromosomes can be distinguished, although genetic differences may sometimes be identified (Graves, 2006). Muller (1914) first suggested that sex chromosome pairs evolved from a pair of autosomes. This evolution can be attributed to suppression of recombination at the sex-determining locus in order to keep the identity of the gonosomes as either an X or a Y and as a result become heteromorphic (Graves, 1998). Additional theory proposed sexual antagonism as one of the driving forces of speciation, where an allele is beneficial in one sex and selected against in another (Rice, 1987). Sexually antagonistic selection can also be accountable for driving the divergence of sex-chromosome systems between closely related species (Van Doorn and Kirkpatrick, 2007). Sex-chromosome turnover has been observed across many fish species, including cichlid fish (Ezaz *et al*., 2006; Mank *et al*., 2006; Kitano *et al*., 2009).

Sex chromosomes demonstrate unique characterizations comparing to gonosomes. Y-specific region is enriched with highly repetitive satellite sequences that were not found in the X sequence (Kondo *et al*., 2006). Furthermore, sex chromosomes were found to be enriched with repetitive genetic elements such as genes, pseudogenes, microRNA, snRNA and tRNA (Kondo *et al*., 2006; Anderson *et al*., 2012). The common characteristic of all known SD genes is their duplication and speciation from autosomal genes. For example, *SRY*, *dmrt1Y* and *amhy* were suggested to have evolved from *SOX3*, *dmrt1* and *amh* duplication in mammals (Foster and Graves, 1994), Medaka (Nanda *et al*., 2002) and Patagonian pejerrey (Hattori *et al*., 2012), respectively. Hence, duplications and copy number variation (CNV) are universal features of SD regions and genes.

Sex Determination in Tilapia

Fish of the genera *Oreochromis*, *Sarotherodon* and *Tilapia* of the family Cichlidae, known as tilapia are among the most important cultured fish in tropical and subtropical countries. Cichlids are a diverse fauna and represent the most species-rich family of vertebrates with more than 3,000 species wildly distributed. Cichlids have shown a capacity for rapid radiation and speciation (Schliewen *et al*., 1994; McKaye *et al*., 2002). Tilapia is the common name for nearly a hundred species and subspecies, and the main cultured species are *Oreochromis niloticus*, *O. mossambicus* and *O. aureus*, *O. urolepis hornorum* and hybrids between them. Since females tend to reproduce at a small size, and to overcome unwanted reproduction, commercial production of tilapia relies on monosex culture of males.

There are several methods available for production of all-male populations in tilapia (Beardmore *et al*., 2001) such as sorting by external sex characteristics, hormonal sex-reversal, inter-specific crosses and mating genetically-modified-tilapia ('YY male'). All proved, so far, difficult to maintain all-male populations sustainably in production environments (reviewed by Cnaani and Levavi-Sivan, 2009). The most commonly used method is sex-reversal using hormonal treatment of sexually undifferentiated fry by administration of hormones or hormone analogues. However, this method is problematic due to environmental and health issues. Additionally, researchers found crosses between different *Oreochromis* species resulting in monosex hybrids, for example, all-male progeny can be produced by *O. niloticus* x *O. u. hornorum* crosses (Wohlfarth *et al*., 1990) and *O. niloticus* x *O. aureus* crosses (Pruginin *et al*., 1975; reviewed by Wohlfarth and Hulata, 1981). A better understanding of the genetic basis of SD in tilapia is needed for overcoming the difficulties in production of all-male populations.

The karyotypes of various tilapia species are highly similar, consisting of 22 chromosomes pairs and morphological differences have not been observed in any chromosome pair as indication of sex chromosomes (Majumdar and McAndrew, 1986; Crosetti *et al*., 1988). Despite the limited differentiation in the putative sex chromosomes, a variety of evidence suggests that SD in tilapias is principally monofactorial (Carrasco *et al*., 1999; Campos-Ramos *et al*., 2001; Griffin *et al*., 2002). The hypothesized sex chromosome systems of the major tilapia species are XX/XY system for *O. mossambicus* and *O. niloticus*; and ZZ/WZ system for *O. aureus* and *O. u. hornorum*. These different sex chromosome systems (XX/XY and ZZ/WZ) have apparently evolved independently (Ayling and Griffin, 2003; Charlesworth *et al*., 2005), due to the fact that different genes are responsible for determining sex in the two sex chromosome systems.

The primary support for these hypotheses comes from breeding hormonally sex-reversed animals. For instance, crosses of sex reversed (ΔXX) males with normal (XX) females produced only females (Mair *et al*., 1991). However, variation in progeny sex-ratios (i.e. other than 1:1) among *O. niloticus* females, suggests that other factors may alter the major genetic sex determiner (Mair *et al*., 1991; Tuan *et al*., 1999). Even though genetic influence on sex is highly heritable, it does not seem to be able to completely override the external environmental influences determining the phenotype (Lester *et al*., 1989). Therefore, sex ratios among *O. niloticus* are easily distorted from the 1:1 primary sex ratio.

Additionally, hormonally sex-reversed *O. aureus* females (ΔZZ) crossed with genetically normal males (ZZ) produced all-male progeny, although some females were occasionally observed (Lahav, 1993; Rosenstein and Hulata, 1994). Gynogenesis resulted in a female-biased sex ratio among the progeny, consistent with the idea that WW and WZ genotypes yield females, while ZZ genotypes develop as males (Avtalion and Don, 1990). Mair *et al*. (1991) concluded that this species has a female (WZ) SD system coupled with additional autosomal influences. They suggested that female-biased sex ratios were produced by unusual male (WZ) genotypes leading to 3F:1M sex ratio (Mair *et al*., 1991). Analysis of sex-ratio in *O. niloticus* x *O. aureus* F_2 and back-crosses indicated their high variability and involvement of additional factors, defined as autosomal factors. Based on this analysis and the various SD regions detected, an autosomal theory was proposed (Hammerman and Avtalion, 1979). According to this theory, *O. niloticus* males and females were defined as aaXY and aaXX, respectively, with emphasis on aa autosomes that do not segregate and have no influence on SD in *O. niloticus*. However, *O. aureus* males and females were defined as AAZZ and AAZY, respectively, showing potential appearance of new quantitative trait loci (QTL) for SD through hybridization (Hammerman and Avtalion, 1979).

External factors such as temperature can override genetic factors that determine sex in tilapia, thereby skewing sex ratios (Baroiller and D'Cotta, 2009). At high temperatures (above 32°C), the level of aromatase mRNA and estradiol decreases, resulting in male-biased sex ratios in *O. niloticus* (D'Cotta *et al*., 2001).

Although genetic and environmental SD mechanisms have long been thought as distinct factors, recent data show that the integration of both factors ultimately guides the bipotential gonad towards the male or the female fate (Barske and Capel, 2008; reviewed in Baroiller *et al*., 2009). This is clearly exemplified in tilapia, which has a ZZ/WZ system as well as XX/XY system in other strains, in which the sex ratio can be modified by temperature, and moreover, in which autosomal chromosomes can influence the definitive sex of the fish (reviewed in Devlin and Nagahama, 2002).

Genetic Infrastructures Available in Tilapia

The first genetic map of *O. niloticus* spanned 704 cM in 30 linkage groups covering the 22 chromosomes of this species (Kocher *et al*., 1998) with 162 DNA markers (microsatellite and AFLP). Lee at al. (2005) published the second generation linkage map that spanned 1,311 cM in 24 linkage groups containing 525 microsatellite and 21 gene-based markers. This map enabled researchers to identify genomic regions associated with genes affecting various traits, e.g. growth, salinity and cold tolerance with segregating families of tilapia (Cnaani *et al*., 2004; Korol *et al*., 2007).

The physical map of *O. niloticus* was released by the Broad Institute in 2011 (NCBI Assembly GCA_000188235.2). This infrastructure has set the base for application of genomic tools such as next generation sequencing and microarrays. For example, sequencing of total RNA extracted from different tissues/treatments, followed by mapping the sequences to the genome, enable the analysis of genes' abundance representing gene expression. The combination of linkage and physical maps allows fine mapping of detected QTLs and the genes harbored in their regions. Lee *et al*. (2010) published 116,899 expressed sequence tags (ESTs) from 17 normalized and two non-normalized cDNA libraries representing 16 tissues from tilapia. This data was an important resource for analysis of gene expression and annotation of genome sequences.

Additional infrastructure of Nile tilapia is the ability to produce monosex families after genetic and hormonal manipulations. GMT (genetically modified tilapia) are fish that result in progeny average sex ratio of >95% male, also known as the "YY male technology" (Fishgen Ltd). All-female progeny are produced by mating with sex-reversed males (females fed with testosterone during early period of life). For SD research, the ability to control and predetermine the sex of progeny allows analysis of gene expression in males *vs.* females at an early age or even at the embryo level when visual sex calling by gonad squash technique is not possible (Mair *et al*., 1997).

Detection of Sex Determining QTL in Tilapia

The differences in the SD mechanisms among closely related tilapia species and the probable influence of both sex-determining genes and the environment, suggest that SD should be treated as a quantitative trait, and analyzed using a marker-based QTL approach (Baroiller *et al*., 2009). Mapping SD QTL in tilapias is based on the second-generation genetic linkage map of tilapia published by Lee *et al*. (2005). Cnaani *et al*. (2008) summarized microsatellite DNA markers associated with gender in eight tilapia strains and species on LG1 and LG3. Additional weak association of a QTL region to SD was reported on LG23 (Cnaani *et al*., 2004). Lee *et al*. (2005) classified the *O. niloticus* as having a male heterogametic (XY) SD system linked to LG1. However, variation in progeny sex-

ratios (i.e. other than 1:1) in *O. niloticus* suggests that other factors may influence the major genetic sex determiner (Mair *et al*., 1991; Tuan *et al*., 1999). Genetic markers on *O. aureus* associated with gender were detected on LG3 and LG1 (Lee *et al*., 2004) suggesting that a female heterogametic (WZ) SD locus is localized on LG3, while a male heterogametic (XY) SD locus is localized on LG1. It was also proposed that SD is based on epistatic interaction, where the dominant female determiner, W on LG3 suppresses the male determiner, Y on LG1. Thus, males are observed only in the absence of the dominant female determiner (Lee *et al*., 2004). Additionally, sex-specific mortality could mimic QTL for SD through elimination of males or females at embryonic or adult developmental period (Shirak *et al*., 2002). Deleterious effects of three marker haplotypes were found on LG2 and LG6 (Palti *et al*., 2002) that are associated with sex-ratio distortions (Shirak *et al*., 2002). Furthermore, *amh* and *dmrt2a2* genes were mapped to LG23 (Shirak *et al*., 2006) adjacent to two previously-reported QTL associated with SD (Cnaani *et al*., 2003, 2004).

Candidate Genes Approach for Sex Determination

Sexual differentiation is generally assumed to be initiated by a major gene, a sex locus, and then a multitude of other genes that follow in a domino effect. Recent studies revealed that the major genes involved in the SD pathway are common to mammals and fish (Schartl, 2004). Moreover, the upstream genes on the SD cascade may vary among organisms whereas the downstream components tend to be conserved (Chelsworth and Mank, 2010). The major genes of the mammalian SD pathway, including *SOX9, WT1, DAX1, SF1, OTCYP19, WNT4, DMRT1*, and *AMH*, have been detected in many vertebrate species and were further analyzed as putative candidate genes (Schartl, 2004; Rodriguez-Mari *et al*., 2005; Birk *et al*., 2000; Ijiri *et al*., 2008). Shirak *et al*. (2006) mapped 11 genes: *amh, cyp19, dax1, dmrt2, dmrta2, fhl3, foxl2, ixl, lhx9, sf1*, and *sox8* to the tilapia linkage map, and located *amh* and *dmrta2* within two distinct QTL regions of LG23, for SD and sex-specific mortality, respectively. *Amh* is considered as the main candidate gene for SD in *O. niloticus* but not in other tilapia species e.g., *O. aureus*. Moreover, genes participating in pathways that are not associated with SD are not considered as candidate genes. Hence, it is important to conduct a systematic genome-wide search for genes involved in SD.

Time Window for Sex Determination in Tilapia

The critical period of sensitivity for SD for Nile tilapia was determined from fertilization to 21 days post hatching (dph) after administration of different androgens, estrogens, or precursor steroids through immersion or diet experiments (Devlin and Nagahama, 2002). In a more recent study, Ijiri *et al*. (2008) showed that differential expression of genes occurring in XX and XY gonads during the

period of 5–6 dph is critical for differentiation of gonads into either ovary or testis in Nile tilapia. This embryonic stage is equivalent to 9-10 days post fertilization (dpf). Rougeot *et al*. (2008) applied temperature treatment on a population of all-female embryos until hatching and showed ~20% phenotypic change in sex of females to males.

In addition, the findings of sex-specific mortality closely after hatching (Palti *et al*., 2002) indicates that the initiation of SD pathways begins in the first few days of embryonic development and ends a month later. These publications demonstrate occurrence of SD pathways near fertilization but no previous work studied differential expression between genders at whole embryo before bi-potential gonad can be detected. Thus, in the current study we focus on gene expression and regulation in the early developing embryo from 2 to 9 dpf.

The Role of MicroRNA in Sex Determination

MicroRNAs (miRNAs) are small noncoding RNAs, about 21 nucleotides in length that may regulate up to 30% of gene expression by partial or complete pairing with complementary sequences of its target messenger RNAs (mRNA), thus controlling the protein production (Berezikov, 2011). miRNAs are highly conserved (Lewis *et al*., 2003; Li *et al*., 2010) among eukaryotes and considered important elements in many biological processes during development, such as cell growth, differentiation and apoptosis (Berezikov, 2011). Due to miRNAs abundance and diversity, they become key elements, both in the speciation process and as phylogenetic markers.

The vast majority of miRNAs found in vertebrate genomes needs to be functionally characterized. Identifying tissue-specific miRNAs is the first step toward understanding the biological functions of miRNAs, which include the regulation of tissue differentiation and the maintenance of tissue identity. Giraldez *et al*. (2005) reported on the essential role of miRNAs for development and adult function of all tissues and for cell fate determination, axis formation, and cell differentiation morphogenesis in zebrafish embryos. These findings suggest that miRNAs can play a significant role in development and more specifically in SD. Profile expression of chicken miRNAs around gonadal sex differentiation identified sexually dimorphic miRNAs and expression patterns. Expression of gga-miR-202-5p was observed to be sexually dimorphic, with up-regulation in the developing testis from the onset of sexual differentiation (Bannister *et al*., 2009). Studies on mouse characterized 55 miRNAs differentially expressed in testis and ovary (Mishima *et al*., 2008) and illustrated their importance for the proliferation of PGCs and spermatogonia (Hayashi *et al*., 2008). Other than in zebrafish, miRNAs composition and expression in fish was hardly investigated. Considering their evolutionary conservation and regulatory functions, they probably play a role in tilapia SD regulation.

The goal of this thesis was to fine map previously detected QTL with our segregating families and the genes harbored in their regions, and to identify genes and miRNAs involved in SD pathways in Nile tilapia.

Research objectives

1. Linkage mapping - Testing of known QTLs in tilapia for their effects on sex determination.
2. Physical mapping - Fine-mapping of the major QTL for SD on LG23.
3. Functional mapping - Detection of genes and microRNAs with differential expression between genders at early embryonic development.

Part A – Linkage mapping

ANIMAL GENETICS Immunogenetics, Molecular Genetics and Functional Genomics

SHORT COMMUNICATION doi:10.1111/j.1365-2052.2010.02128.x

Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*)

O. Eshel*,†, A. Shirak†, J. I. Weller†, T. Slossman†, G. Hulata†, A. Cnaani† and M. Ron†

*The Hebrew University of Jerusalem, Robert H. Smith Faculty of Agriculture, Food and Environment, P.O. Box 12, Rehovot, 76100, Israel.
†Institute of Animal Science, Agricultural Research Organization, P.O. Box 6, Bet Dagan, 50250, Israel

Summary

Genetic markers in tilapia species associated with loci affecting sex determination (SD), sex-specific mortality or both were mapped to linkage groups (LG) 1, 2, 3, 6 and 23. The objective of this study was to use these markers to fine-map the locus with the greatest effect on SD in *Oreochromis niloticus*. Our parental stock, full-sibs of Nile tilapia (Swansea origin), were divided into three groups: (i) untreated, (ii) feminized by diethylstilbestrol and (iii) masculinized by 17α-methyltestosterone. We analysed the first group for association of microsatellite markers representing these five LGs. The strongest association with gender was found on LG23 for marker *UNH898* (χ^2; $P = 8.6 \times 10^{-5}$). Allele 276 was found almost exclusively in males, and we hypothesized that this allele is a male-associated allele (MAA). Sex-reversed individuals were used for mating experiments with and without the segregating MAA. Mating of individuals lacking the MAA resulted in all-female progeny. Mating of two heterozygotes for MAA gave rise to 81 males and 30 females. Analysis of association between gender and genotypes identified the MAA in 98.6% of males as opposed to 8.0% of females (χ^2; $P = 2.5 \times 10^{-18}$). Eight markers that flank *UNH898* were genotyped to map the locus on LG23 within a confidence interval of 16–21 cM. Mating of homozygous individuals for MAA is underway for production of all-male populations.

Keywords locus, monosex progeny, *Oreochromis niloticus*, sex determination, sex ratio, tilapia.

Commercial production of tilapia is based on monosex cultures of males (Beardmore *et al.* 2001). To overcome the difficulties for production of all-male cultures, a better understanding of the genetic basis of sex determination (SD) in tilapia is needed. The hypothesized sex chromosome mechanism in Nile tilapia (*Oreochromis niloticus*) is an XX/XY system, and evidence for this is based on primary support from breeding hormonally sex-reversed gynogenetic animals (Mair *et al.* 1991). Differences in the SD mechanism among closely related tilapia species and sensitivity to different environmental factors suggest that gender could be analysed as a multifactorial trait (Baroiller *et al.* 2009).

Previous reports have shown association between microsatellite DNA markers and gender on linkage groups (LG) 1, 3 and 23 because of elimination of specific genotypes in one of the two genders (Cnaani *et al.* 2004; Lee *et al.* 2004). In addition, sex-specific mortality was reported for markers on LG 2, 6 and 23 (Shirak *et al.* 2002). The aim of this study was to utilize previously reported sex-associated markers on LG 1, 2, 3, 6 and 23 to fine-map the locus with the greatest effect on SD in Nile tilapia.

Full-sib fry of Nile tilapia (Swansea stock) was divided into three groups: (i) untreated, (ii) feminized by diethylstilbestrol and (iii) masculinized by 17α-methyltestosterone. Hormone treatments were given orally by feeding the fry during the four-first weeks. At an age of 3–4 months, the untreated fish were phenotyped for sex, and fin samples were taken for DNA extraction. Initially, 25 females and 20 males of the untreated group were genotyped for microsatellites *BYL002* (HQ199073), *UNH159*, *GM139*, *GM390* and *UNH898*, located on LG 1, 2, 3, 6 and 23, respectively. A goodness-of-fit test (Pearson's χ^2) was conducted to evaluate the association between gender and genotypes of the microsatellite markers. Significant association with gender was found for the markers on LG 1 ($P = 0.01$) and 23 ($P = 8.6 \times 10^{-5}$). We focused on the most significant locus on LG23. Analysis of the untreated individuals for *UNH898* indicated that allele 276 was found almost exclusively in males (1♀: 52♂). Thus, we hypothesized that this allele is a

Address for correspondence

M. Ron, Department of Cattle and Genetics, Institute of Animal Science, Agricultural Research Organization, P.O. Box 6, Bet Dagan, 50250, Israel.
E-mail: micha@agri.huji.ac.il

Accepted for publication 29 June 2010

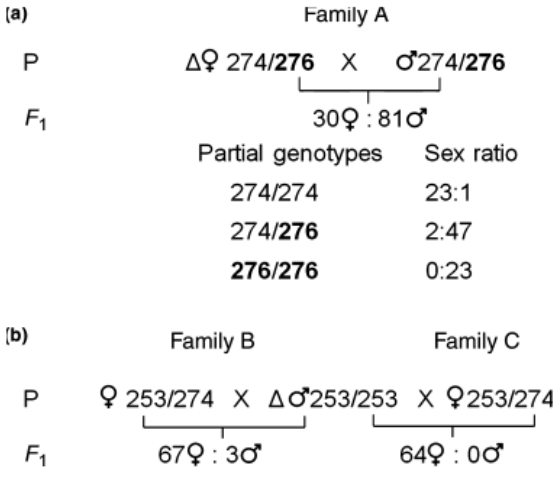


Figure 1 Sex ratio and partial genotypes for *UNH898* in progeny of parents segregating (a) and non-segregating (b) for the male-associated allele (MAA). 253, 274 and 276 are fragment lengths of alleles in bp. The MAA is in bold face; Δ denotes sex-reversed individual.

male-associated allele (MAA). DES-treated females with and without MAA were classified as sex-reversed females and genetically females, respectively. Likewise, the MT-treated males with and without MAA were classified as genetically males and sex-reversed males, respectively. We designed two mating experiments with and without the segregating MAA (Fig. 1).

In family A, the cross between two segregating parents for MAA yielded a male to female ratio that is not significantly different from the expected 3:1 ratio. Analysis of association between gender and genotypes identified the MAA in 98.6% of males as opposed to 8.0% of females (χ^2; $P = 2.5 \times 10^{-18}$). In families B and C, the crosses between non-segregating parents for MAA yielded progeny that consisted of 96 and 100% females, respectively. Thus, the hypothesis that the MAA has a major role in SD was supported by the outcome of these experiments.

To fine-map the locus on LG23, we genotyped eight markers that flank *UNH898* using individuals of the untreated group and family A. The linkage map was computed with the 'fixed' option of CRIMAP (http://linkage.rockefeller.edu/soft/crimap/). The markers *GM338*, *GM597*, *GM212*, *GM047*, *GM283*, *UNH898*, *UNH216*, *GM631* and *UNH879* were mapped to 16, 16, 16, 17, 17, 18, 26, 33 and 33 cM on LG23, respectively. No other informative markers are known in this region. The interval mapping was based on non-linear regression using the method of Knott *et al.* (1996), with the programme developed by Spelman *et al.* (1996). The test statistic and locus effects were evaluated at one-cM intervals. The 95% confidence interval (CI) for the sex determination locus was determined by the generation of 200 bootstrap samples. The

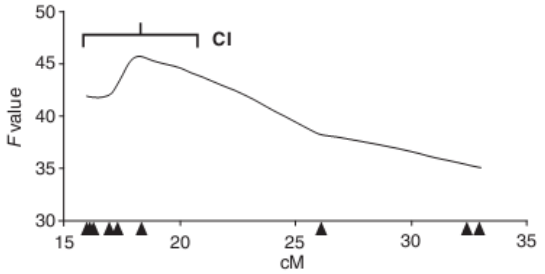


Figure 2 Mapping of the locus for sex determination on LG23. Microsatellites are indicated with ▲ and located from left to right: *GM338*, *GM597*, *GM212*, *GM047*, *GM283*, *UNH898*, *UNH216*, *GM631* and *UNH879*.

maximum *F*-value of the locus on LG23 for SD was 42 at 18 cM, with a CI of 16–21 (Fig. 2).

The CI region harbours two genes, *SOX14* (encoding sex-determining region Y-box 14) and *AMH* (encoding anti-Mullerian hormone), which were previously mapped to this region on LG23 (Shirak *et al.* 2006; Cnaani *et al.* 2007). *AMH* is expressed in Sertoli cells at the onset of sex-differentiation and causes Müllerian duct regression in mammalian males (Baroiller *et al.* 2009). Although the Müllerian duct is absent in tilapia, *AMH* may still play a role in SD and thus can be considered to be a major candidate gene. In the F_1 generation, six of 230 individuals were not in accordance with our hypothesis of a segregating MAA. Thus, the locus on LG23 explains 97.4% of SD (224/230). The incomplete association of MAA with gender may be because of a minor genetic effect found on LG1 and may also be due to double recombination between *UNH898* and the sex-determining gene. Nevertheless, mating of homozygous individuals for MAA is underway for production of all-male populations.

Acknowledgements

This is a contribution from the ARO, Institute of Animal Science, Bet Dagan, Israel, No. 566/10. The research was supported by research grant IS-3995-07 from the United States-Israel Bi-national Agricultural Research and Development (BARD) Fund.

References

Baroiller J.F., D'Cotta H. & Saillant E. (2009) Environmental effects on fish sex determination and differentiation. *Sexual Development* **3**, 118–35.

Beardmore J.A., Mair G.C. & Lewis R.I. (2001) Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* **197**, 283–301.

Cnaani A., Zilberman N., Tinman S., Hulata G. & Ron M. (2004) Genome-scan analysis for quantitative trait loci in an F_2 tilapia hybrid. *Molecular Genetics and Genomics* **272**, 162–72.

Cnaani A., Lee B.-Y., Ozouf-Costaz C., Bonillo C., Baroiller J.F., D'Cotta H. & Kocher T.D. (2007) Mapping of *SOX2* and *SOX14* in tilapia (*Oreochromis* spp.). *Sexual Development* **1**, 207–10.

Knott S.A., Elsen J.M. & Haley C.S. (1996) Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. *Theoretical and Applied Genetics* **93**, 71–80.

Lee B.-Y., Hulata G. & Kocher T.D. (2004) Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* **92**, 543–9.

Mair G.C., Scott A.G., Penman D.J., Beardmore J.A. & Skibinski D.O.F. (1991) Sex determination in the genus *Oreochromis*.1. Sex reversal, gynogenesis and triploidy in *Oreochromis niloticus*. *Theoretical and Applied Genetics* **82**, 144–52.

Shirak A., Palti Y., Cnaani A., Korol A., Hulata G., Ron M. & Avtalion R.R. (2002) Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). *Journal of Heredity* **93**, 270–6.

Shirak A., Seroussi E., Cnaani A., Howe A.E., Domokhovsky R., Zilberman N., Kocher T.D., Hulata G. & Ron M. (2006) *AMH* and *DMRTA2* genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* **174**, 1573–81.

Spelman R.J., Coppieters W., Karim L., van Arendonk J.A.M. & Bovenhuis H. (1996) Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. *Genetics* **144**, 1799–807.

Part B - Physical mapping

Linkage and Physical Mapping of Sex Region on LG23 of Nile Tilapia (*Oreochromis niloticus*)

O. Eshel,*,† A. Shirak,† J. I. Weller,† G. Hulata,† and M. Ron†,1
*Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot 76100, Israel, and †Institute of Animal Science, Agricultural Research Organization, Bet Dagan 50250, Israel

ABSTRACT Evidence supports that sex determination (SD) in tilapia is controlled by major genetic factors that may interact with minor genetic as well as environmental factors, thus implying that SD should be analyzed as a quantitative trait. Quantitative trait loci (QTL) for SD in *Oreochromis niloticus* were previously detected on linkage groups (LG) 1 and 23. Twenty-one short single repeats (SSR) of >12 TGs and one single nucleotide polymorphism were identified using the unpublished tilapia genome sequence on LG23. All markers showed two segregating alleles in a mapping family that was obtained by a cross between *O. niloticus* male (XY) and sex-reversed female (ΔXY) yielding 29 females (XX) and 61 males (XY and YY). Interval mapping analysis mapped the QTL peak between SSR markers ARO172 and ARO177 with a maximum F value of 78.7 ($P < 7.6 \times 10^{-14}$). Twelve adjacent markers found in this region were homozygous in females and either homozygous for the alternative allele or heterozygous in males. This segment was defined as the sex region (SR). The SR encompasses 1.5 Mbp on a single tilapia scaffold (no. 101) harboring 51 annotated genes. Among 10 candidate genes for SD that were tested for gene expression, anti-Müllerian hormone (*Amh*), which is located in the center of the SR, showed the highest overexpression in male *vs.* female embryos at 3 to 7 days postfertilization.

KEYWORDS
sex region
linkage mapping
physical mapping
Oreochromis niloticus
microsatellite markers

Sex determination (SD) can be controlled by one or more genetic factors, environment or their interactions, involved SD factors located on sex chromosomes and/or on either autosomes (Bull 1981). The sex chromosomes are characterized by both morphologically undifferentiated and differentiated homologs, in simple and multiple systems with male or female heterogamety. Studies on organisms with differentiated sex chromosomes, male heterogametic (mammals and fly) and female heterogametic (birds and reptiles), have shown interesting similarities between the two systems (XY/XX and ZW/ZZ) of sex determination (Ellegren 2011).

Teleost species are an interesting model for SD research, with a variety of SD systems and capability of producing viable hybrids between closely related species having different SD systems (Mank *et al.* 2006). Teleost fish have diverged from land vertebrates more than 450 million years ago, after they diverged from birds and mammals, but before they diverged from each other (Bellott *et al.* 2010). Moreover, these fish species experienced whole-genome duplication (ancestral tetraploidy), followed by variable-rate reduction of ploidy that significantly complicates the identification of orthologs (Kasahara *et al.* 2007). Bellott *et al.* (2010) compared zebrafish, Tetraodon, pufferfish, and medaka genomes to mammalian X or avian Z chromosome and reported that most orthologs to Z and X genes occupy separate portions of each fish genome.

Different aspects of tilapia SD have been explored because tilapias are an important aquaculture commodity (Devlin and Nagahama 2002). Their commercial production relies on all-male monosex culture, which so far has proved difficult to maintain in large-scale production facilities (Cnaani and Levavi-Sivan 2009). A better understanding of the genetic basis of SD in tilapia is needed to overcome these difficulties.

The differences in SD mechanisms among closely related tilapia species and the influence of the environment (Baroiller *et al.* 2009) suggest that SD should be analyzed as a quantitative trait using a markers-based QTL approach. Various sex-linked markers have been identified in *O. niloticus* and *O. aureus* (Lee *et al.* 2003, 2004; Shirak *et al.* 2002, 2006; Eshel *et al.* 2010) and mapped to different LG. In purebred *O. niloticus* and *O. niloticus* × *O. aureus* hybrids, the QTL

doi: 10.1534/g3.111.001545
Manuscript received July 28, 2011; accepted for publication November 3, 2011

Supporting information is available online at http://www.g3journal.org/lookup/suppl/doi:10.1534/g3.111.001545/-/DC1
[1]Corresponding author: Institute of Animal Science, Agricultural Research Organization, P. O. Box 6, Bet Dagan 50250, Israel. E-mail: micha@agri.huji.ac.il

were detected on LG1, LG23 (Lee *et al.* 2003; Eshel *et al.* 2010) and on LG3 (Lee *et al.* 2005), respectively. The SD QTL on LG23 was mapped within a confidence interval (CI) of 16–21 cM (Eshel *et al.* 2010; Figure 1B), which harbors the genes *Amh* and *Dmrta2* that are involved in the vertebrate SD cascade (Shirak *et al.* 2006, Figure 1A). The first assembly version of the unpublished tilapia genome, consisting of 5900 scaffolds, was recently released (Accession no. PRJNA59571). Using this information enabled us to refine the confidence interval of SD QTL on LG23 and find positional candidate genes for the sex master-key regulators in the Swansea stock of *O. niloticus*.

The critical period for SD in tilapias is 0–18 days postfertilization (dpf). During this period, embryos are sensitive to androgens, estrogens, and precursors of steroids through immersion and dietary exposure (Devlin and Nagahama 2002). In a more recent study, Ijiri *et al.* (2008) demonstrated that differential expression of genes in XX and XY gonads of *O. niloticus* during the period of 9–10 dpf is critical for differentiation of primordial germ cells (PGC) into either ovary or testis. Rougeot *et al.* (2008) applied temperature treatment on all-female population embryos until hatching (3–4 dpf) and showed ~20% phenotypic sex change of females to males. Furthermore, Palti *et al.* (2002) demonstrated by using genetic markers that sex-specific mortality occurs shortly after hatching. On the basis of these findings, we hypothesized that master regulation genes initiating the SD cascade should be expressed before the detectable differences in the PGCs.

Recent studies revealed that the major genes involved in the SD pathway are common to mammals and fish (Schartl 2004). Moreover, the upstream genes on the SD cascade may vary among organisms, but downstream components tend to be conserved (Charlesworth and Mank 2010). To study the onset of the SD cascade at early stages in embryonic development, we selected eight genes (*Lhx9*, *Amh*, *Foxl2*, *Cyp19a*, *Dmrt1*, *DAX1*, *Sox9a*, and *Sox9b*) with a known role in the SD pathway of other organisms (Birk *et al.* 2000; Shirak *et al.* 2006; Ijiri *et al.* 2008) and examined their expression from gastrula to late larval stage (2–9 dpf) in *O. niloticus* monosex progeny of XY males and XX females. Additionally, we analyzed two genes previously mapped to the SD region: *Sox14* (Cnaani *et al.* 2007) and *ELAVL1* (A. Shirak, unpublished data) with ambiguous similarity to known genes on the SD pathway. Gene expression profiles represent the

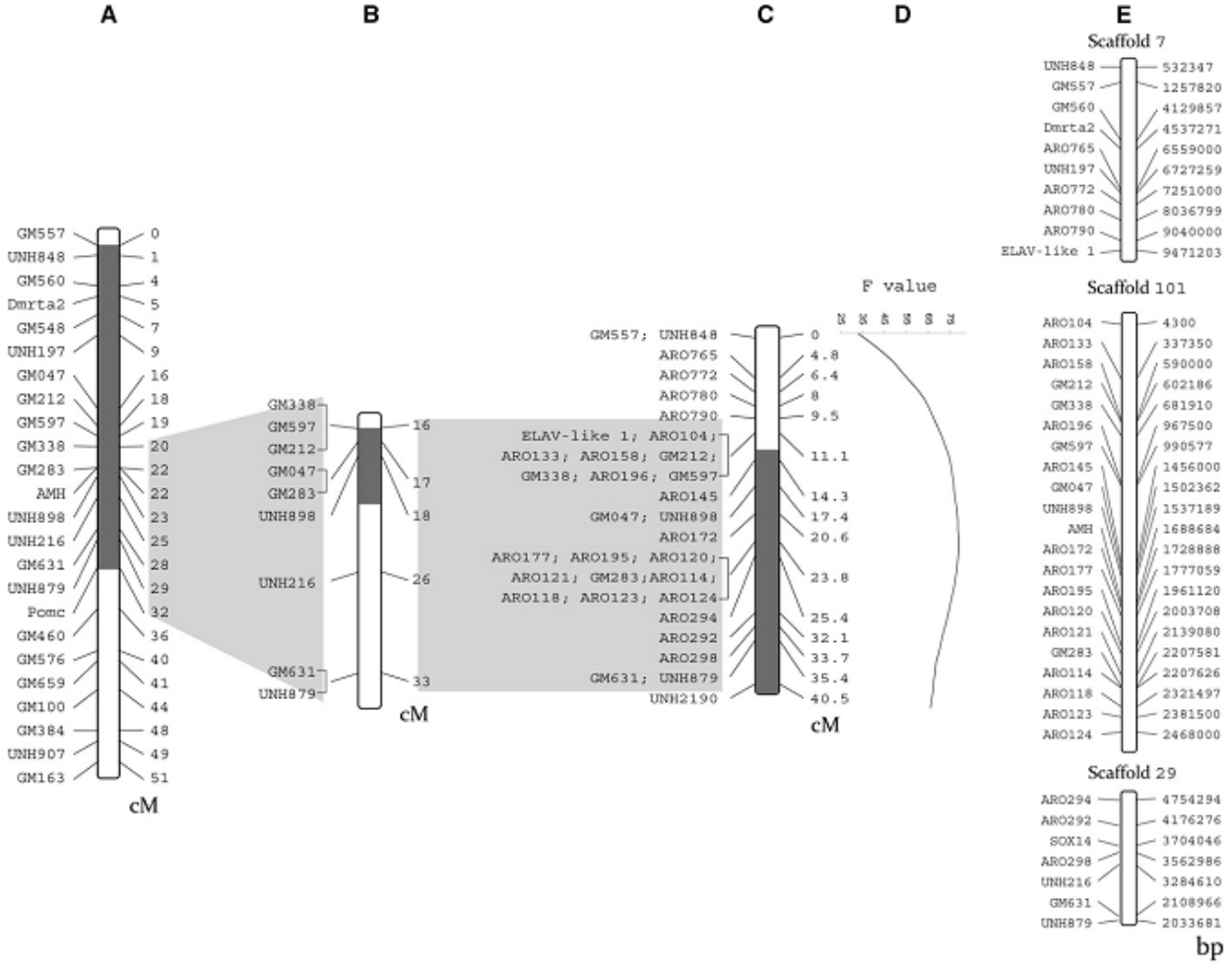


Figure 1 Linkage and physical map of SR on LG23 and interval mapping of the SD QTL. (A) LG23 genetic map by genotyping of *O. niloticus* ♂ × *O. aureus* ♀ family (Shirak *et al.* 2006). (B) Sex determination QTL on LG23 for *O. niloticus* families based on nine SSR markers (Eshel *et al.* 2010). (C) Fine mapping of the QTL region with additional mapping family (sex reversed) with 33 markers. (D) Interval mapping of the QTL based on panel C. (E) LG23-related scaffolds. QTL, quantitative trait loci; SD, sex determination; SR, sex region; SSR, short single repeat.

Table 1 Primers for SSRs and *ELAV-like 1* gene used in this study and their locations on the unpublished tilapia genome scaffolds

Scaffold	Marker/Gene	Forward Primer	Reverse Primer
7	ARO765	CCTGAAACTCAGGCGCTGTA	GCTCTCACCAAGGTCAGCAA
	ARO772	GCCTTGTGCCACTGTAGGAG	AACCTGCCTCTCCTGGAATC
	ARO780	TGTGGGGTTTTTGAAGCCTA	GAAACCCCCTCTTCCTTGTG
	ARO790	TGAAGCAAACAGAGGCCATT	GCTGGGTGAGGGGTTTTGTA
	ELAV-like 1	GCTTTGATAAGAGGGCTGAGG	AGTTCCTGGCCTGGTTGG
		Extension primer: [AAAAAA]CAAACACCTGAACGGACACAC[a]	
29	ARO298	CAGACTGTCCCCATCCTCAA	AGGGAGCTGGATCTGCCTAA
	ARO292	TTGACTACCGGCTTGCATTC	GCCCGAACATAAGATGTCCA
	ARO294	TGCTCTCACTGCTGAGCAAA	CGCAAATGTTAGGCCAGAAA
101	ARO104	AAGACCCGTTCTTCGTCGTC	TTCATTCCACCTGCTCCAAA
	ARO133	GTGAGGCAAGTCCGGTTTCT	TGATCCACGGCGTATTGAGT
	ARO158	GTGGGCAAAAACAAGCCATT	TGTTTCAGTGTGAACGTGTGTG
	ARO196	GATTGTGGCCTGGTCAAGTG	TCCGTTTGTCTGCTGTGTGA
	ARO145	CAATGTGGCAATGTGTCCAA	CGGTGTCTCTGTGTCGTGTG
	ARO172	AGGCCTTTCATCGCTGTTTT	ACCCTGTAGATGAGCGCAAA
	ARO177	CCCTGCCCTGAACTACCTTC	GCTGCAAGCAAATGAAAAGC
	ARO195	CATGCTGATGGAGACCGATT	TCAAGACGCAATGGAGTGTG
	ARO120	AAGGGAAAGTGGCTCAGCTC	GTTGCTTCCCCACAGTTTCA
	ARO121	GGTGGGACTGTGGTGTATGG	GGTGGATTGCAAGCAACATT
	ARO114	AGGAGAAGTCGCAGGTGACA	GGCACAGTTGCCTGGTACAT
	ARO118	TGAATCTTCCCACAGCAACA	GTTGGTGCCAACAAAGCAAT
	ARO123	TTAATCCTGCCCACCTCTCC	AAGCAAAAGCATTTTCATGTTCA
	ARO124	CGAGCTGCTTTGTTGTCTGA	CGAACCGAAAATGAGAATGC

SSRs, short single repeats.
[a] [AAAAAA] is a stabilizing tail.

primary level of integration between environmental factors and the genome, providing the basis for phenotypes, such as morphology and behavior. Therefore, we examined differences in candidate genes expression between genders during early embryonic development for initiation of the SD cascade in tilapia.

The objectives of this study were to refine the sex region on LG23 using both linkage and physical mapping and to identify candidate genes for SD in the region with differential expression at early embryonic development.

MATERIALS AND METHODS

Breeding of *O. niloticus* (Swansea stock) families used for this study was performed at the Agricultural Research Organization, Israel.

Mapping family

The inheritance of gender in a cross between *O. niloticus* male (XY) and a sex-reversed neofemale (ΔXY) that yielded 29 females (XX) and 61 males (XY and YY) was validated by segregation of the sex-linked marker *UNH898* (Eshel *et al.* 2010).

Monosex groups

To obtain all-female (XX) and all-male (XY) progenies, eggs of a single *O. niloticus* female (XX) were divided into two groups, and each group was artificially fertilized with either milt of a sex-reversed male (ΔXX) or milt of a genetically modified male (YY). Sex was determined at age of three months by gonadal squash of at least 100 individuals per each full-sib group (Mair *et al.* 1997).

Development of SSR markers: We ran BLASTN search for the *Amh*, *Dmrta2*, and *Sox14* genes and for nine SSR markers that were previously mapped to LG23 (Lee *et al.* 2004; Shirak *et al.* 2006; Cnaani *et al.* 2008) against the unpublished tilapia genome (Accession no. PRJNA59571; http://cichlid.umd.edu/blast/blast.html). Hits were found in three scaffolds: no. 7, 101, and 29. We searched for tandem

Table 2 Primers for genes analyzed by qPCR

Gene	Forward Primer	Reverse Primer
Amh	GCACCCAGCTGCAGTACAC	GTGGGAGGTCAAAGGTCAAC
Cyp19a1a[a]	GCATAGGCACAGCCAGCAAC	GTGCACTGCTGAAGATCTGCTTAGTA
Dax1	CAGATCTGGAGGGTTTGC	GATGGATCAGCCTGACGTG
Dmrt1[a]	CGGCCCAGGTTGCTCTGAG	CCAACTTCATTCTTGACCATCA
ELAV-like 1	CAGGCTTCAGGTCTGTCACG	GTGTCCGTTCAGGTGTTTGA
Foxl2	CACGACCAAGGAGAAAGAGC	TGGCAATGAGAGCGACATAG
Lhx9	GATTACTACAGGTTCTCCGTGCAG	TCAGGTGATACACGGAGTCG
Sox14	TGCTCAAGAAGGACCGTTACG	AAGAGCCCAAAAGAGAGTCCG
Sox9a	GCAAACTTTGGAGATTGCTCA	TCGGGGTGATCCTTCTTATG
Sox9b	GAGAGCATTCAGGTCAGTCACA	TCAGATCAGCTTTGCTGGAG
GAPDH	GGCATCGTGGAAGGTCTCAT	CATTTTACCAGAGGGCCCGT

[a] Ijiri *et al.* (2008).

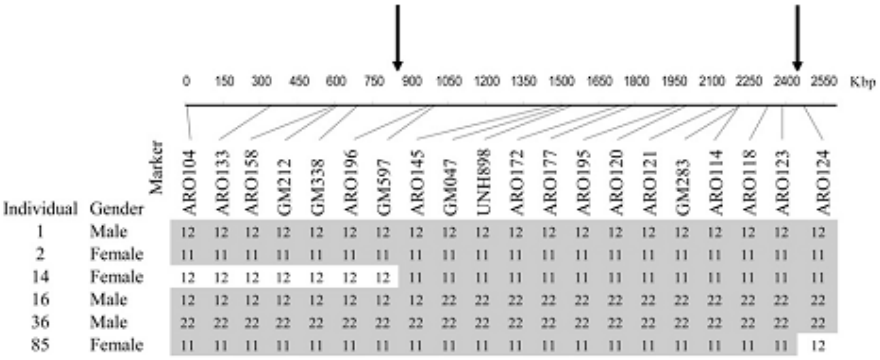


Figure 2 Determination of boundaries of the SR on scaffold 101 based on genotyping data of SSRs for selected individuals. Heterozygous genotypes of females contributed to the reduction of the SR interval delimited by arrows. The two homozygous genotypes are denoted "11" and "22"; the heterozygous is denoted as "12." Females have the 11 genotype for all markers within the SR, whereas males have either 12 or 22. The genotypes' segments corresponding to gender for each individual are denoted with shading. SR, sex region; SSRs, short single repeats.

repeats of >12 TG in scaffold 7 (6,500,000–9,485,422 bp), in scaffold 29 (3,291,196–5,141,938 bp), and in the entire scaffold 101. We entered the sequence of 200 bp upstream and downstream of the TG repeats core to Primer3 software and developed 21 novel SSR markers. Polymorphism of these markers was tested in parents of the mapping family. To develop genetic markers in the vicinity of UNH216, we mapped in our family the marker UNH2190, which was derived from the Malawi cichlids hybrid *Metriaclima zebra* × *Labeotropheus fuelleborni* and was mapped adjacent to UNH216 (Albertson *et al.* 2003).

Development of the SNP marker: On the basis of partial cds sequence (GI: 93115149) of *O. mossambicus ELAVL1* (embryonic lethal, abnormal vision, Drosophila-like 1), we identified the SNP polymorphism A/G (Table 1) at nucleotide 391 in our mapping family.

DNA extraction and genotyping of SSR and SNP markers: DNA was isolated from fin samples by the "salting out" high-throughput procedure (Zilberman *et al.* 2006). The concentration of the DNA was quantified with NanoDrop spectrometer (NanoDrop Technologies, DE), and each DNA sample was diluted to a final concentration of ~10 ng/μl. PCR amplification was performed in a total volume of 10 μl with Super-Therm Taq DNA polymerase (JMR Holding, London), mixture of 2 mM dNTPs of each nucleotide, and primer concentration of 10 pmol/μl (Metabion GmbH, Germany). PCR conditions were 3 min at 94°; 40 sec at 94°, 40 sec at 61.5°, 1 min at 72° for 30 cycles and 10 min at 72°. The mapping family was amplified for SSR markers, and genes with primers taken from NCBI database or designed based on scaffold sequence (detailed list in Table 1), where one primer in each pair was 5′ end-labeled by HEX, TET, or FAM fluorescent dyes (Operon Technologies, Alameda, CA). Size calling of PCR products was determined using ABI GeneMapper software version 4.0 (Applied Biosystems, Foster City, CA) after electrophoresis in a capillary gel on ABI-3130 apparatus. Sequencing and SNaPshot reaction for genotyping of SNP markers were also carried out on ABI-3130 according to the manufacture instructions using the primers specified in Table 1. (For additional data, see supporting information, File S1.)

Linkage and interval mapping: The linkage map for LG23 using segregating markers in our mapping family was reconstructed by CRIMAP software (http://linkage.rockefeller.edu/soft/crimap/).

The interval mapping was based on a nonlinear regression using the method of Knott *et al.* (1996), with the program developed by Spelman *et al.* (1996). The test statistic and locus effects were evaluated at 1 cM intervals. The 95% confidence intervals (CI) for the QTL location and effect were determined by generation of 200 bootstrap samples.

Identification of genes and annotation: Annotation of genes positioned in the SR was performed by combining three bioinformatics resources: (1) EST contigs assembled via MIRA program with BLASTN (Lee *et al.* 2010); (2) comparative mapping to other fish genomes by BLASTX; and (3) BouillaBase annotation using Maker Gene pipeline (http://cichlid.umd.edu/cgi-in/gb2/gbrowse/Tilapia_broad_scaffolds_v1/?source=Tilapia_broad_scaffolds_v1).

Comparative mapping: After determining the boundaries of the SR in LG23, we detected 39 annotated genes in the unpublished tilapia genome database (http://cichlid.umd.edu/cgi-bin/gb2/gbrowse/Tilapia_broad_scaffolds_v1/?source=Tilapia_broad_scaffolds_v1) using Maker pipeline. Further analysis between tilapia and stickleback group VIII (15.4–16.7 Mbp) (http://www.ensembl.org/index.html) identified all 39 genes in the same order in both genomes. On the basis of this high level of orthology, we used stickleback orthologous region on UCSC genome browser (http://genome.ucsc.edu/) as anchor for similarity with Tetraodon (chr1:13.4–14.7 Mbp), medaka (chr4:7.4–8.8 Mbp), fugu (scaffold 25:0.4–1 Mbp), zebrafish (chr22:19.1–21.6 Mbp), and human (chr19p13.3: 1.5–5 Mbp), and we detected 12 additional genes in tilapia.

RNA extraction and qPCR: A pool of 20–30 embryos from each gender were placed in RNAlater reagent (Qiagen) to stabilize the RNA and then stored at −20° until RNA extraction. Total RNA was

Table 3 Level of normalized relative expression ± SD and statistical significance of sex-specific differences for gene candidates for SD in embryos at 2 to 9 dpf

Gene	dpf	All Female	All Male	Probability
ELAVL1	2	14.5 ± 2.2	3.2 ± 2.1	**
	5	42.1 ± 11.8	0.5 ± 11.8	*
	7	22.8 ± 5.8	1.2 ± 5.4	*
	9	2.7 ± 0.6	0.2 ± 0.6	*
Amh	3	0.02 ± 0.005	0.04 ± 0.005	*
	4	0.2 ± 0.2	3.5 ± 0.2	***
	5	0.5 ± 0.9	5.2 ± 0.9	**
	6	0.4 ± 0.2	5.8 ± 0.2	***
	7	0.3 ± 0.7	3.7 ± 0.7	**
Lhx9	2	3.1 ± 0.18	0.8 ± 0.2	***
Sox9a	7	1.6 ± 0.17	0.9 ± 0.2	*
Sox9b	7	6.3 ± 0.6	4 ± 0.6	*
Foxl2	8	0.9 ± 0.06	0.6 ± 0.06	*
Sox14	9	0.3 ± 0.07	0.55 ± 0.06	*

Asterisks represent levels of significance for sex-specific expression differences: *$P \leq 0.05$, **$P \leq 0.01$, and ***$P \leq 0.001$. dpf, days post fertilization.

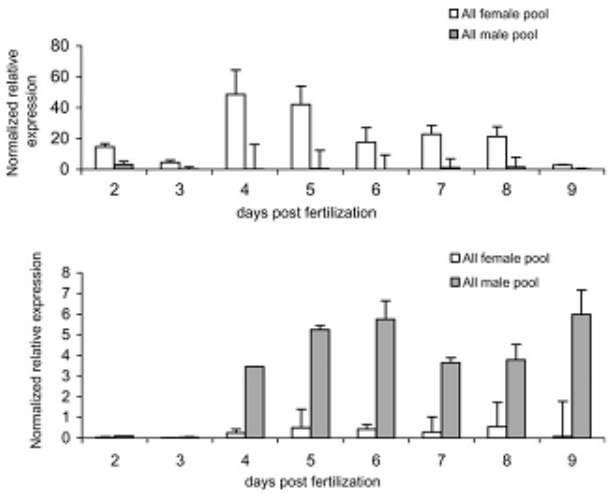


Figure 3 Normalized relative expression of *Amh* (A) and *ELAVL1* (B) for all-male (gray) and all-female (white) pools at 2–9 dpf. Deviation bars represent standard errors and asterisks represent levels of significance for sex-specific expression differences: $^{*}P \leq 0.05$, $^{**}P \leq 0.01$, and $^{***}P \leq 0.001$. dpf, days postfertilization.

extracted (miRNeasy Mini kit, QIAGEN) and analyzed with Agilent 2100 Bioanalyzer (Agilent Technologies, Palo Alto, CA). Synthesis of cDNA was done with SuperScript II (Invitrogen) according to the manufacturer's instructions. Three biological and three technical repeats of qPCR were performed using Absolute Blue SYBR Green Rox mix (Thermo Scientific, WI). The primers were designed so at least one strand was specific to an intron-exon boundary (Table 2). The efficiency of the PCR reaction (linear equation: y = slope + intercept) was measured in triplicate on serial dilutions of the same cDNA sample (pool of RT-RNA samples). Efficiencies (E) of qPCR for each reaction were calculated using the following equation: $E = [10^{(1/\text{slope})}]^{-1}$. Melting-curve analysis was performed for each gene for specificity of qPCR products. The relative amount of the target RNA, called the input amount (IA) according to User Bulletin #2 ABI (PRISM7700 Sequence Detection System, Applied Biosystems), was determined by comparison with the corresponding standard curve for each sample. The IA values were calculated as follows: $IA = [10^{((Ct\ -\ \text{intercept})/\text{slope})}]$, where Ct is the cycle threshold for unknown sample. Each transcript level was normalized by division with the expression values of the *GAPDH* gene, which was used as an internal standard. Final results were analyzed by Student *t*-test. (For additional data, see File S2.)

Gene expression in SD-associated organs: To retrieve available gene expression data for all genes embedded within the SR on LG23 we used the "Gene Atlas" expression data for mammals at BioGPS (Su *et al.* 2004; Wu *et al.* 2009; http://biogps.org/#goto=welcome). Differential expression was determined for individual genes in organs relevant for the SD pathways in tilapia such as brain, testis and pituitary.

RESULTS

Mapping new markers on LG23

Two alleles were found for each of the novel 21 SSR markers in the parents of the mapping family. These markers were designated as ARO markers (Table 1) that were physically mapped to scaffolds 7, 101, and 29 and linkage mapped to an interval of 30 cM of LG23 (Figure 1C).

Linkage and physical mapping of the QTL on LG23

In Figure 1, the QTL interval mapping for SD on LG23 is presented based on the reference mapping family (Figure 1A) and *O. niloticus* families (Figure 1B). In the current study, 33 genetic markers were analyzed, including the new SSR markers added (Figure 1C). Interval mapping analysis mapped the SD QTL region to 13–40 cM with a maximum F value of 78.7 ($P < 7.6 \times 10^{-14}$) at 22 cM (Figure 1D). This region was localized to scaffolds 7, 101, and 29. Physical mapping of the scaffolds with the newly developed markers narrowed down the SR to scaffold 101 between markers GM597 and ARO124 from 990,577 to 2,468,000 bp (Figures 1E and 2). The *Amh* gene is located between these markers. The scaffolds relating to LG23 and the physical map of markers are given in Figure 1E based on the unpublished tilapia genome sequence.

The SR on scaffold 101 was inferred from genotypes for SSRs of selected individuals (Figure 2); 12 adjacent markers found in this region were homozygous in females and either homozygous for the alternative allele or heterozygous in males. This segment was defined as sex region. Markers flanking the SR were heterozygous in females, thus reducing the SR interval to 1.5 Mbp between GM597 and ARO124. The boundaries of the SR are marked by arrows.

Gene expression at early developmental stages

Expression analysis of 10 SD-related genes and two genes mapped within the SD QTL on LG23 was performed for embryos of known type (XX or XY) during 2–9 dpf. No significant differences between genders were found for cyp*19a*, *Dmrt1*, or *Dax1*. Significant sex-differential expression was detected for the remaining 7 genes as presented in Table 3. Figure 3A presents the continued elevation in significance for gender-specific differences for *Amh* expression from 3 dpf ($P = 0.03$) to 7 dpf ($P \leq 0.01$). The Y axis indicates normalized expression values, whereas each bar along the X axis indicates a sample. This gene showed the highest sex-differential expression among all 10 tested genes. Significant sex difference was found for *Lhx9* expression ($P = 0.0002$) at 2 dpf, equivalent to the developmental stage of segmentation (Fujimura and Okada 2007) but not later in the embryonic development. Likewise, sex-differential expression was detected for *ELAVL1* ($P < 0.01$) at age of 2 dpf but was attenuated at 5, 7, and 9 dpf ($P \leq 0.05$) (Figure 3B). Gender-specific expression differences for the other 4 genes (*Foxl2*, *Sox9a*, *Sox9b*, and *Sox14*) were detected at later developmental stages (6–9 pdf).

Characterization of genes positioned within the SR

Fifty-one genes were identified within the SR and are presented in Table 4. Thirty-nine genes had expression data in a variety of 91 tissues of mammals in the BioGPS database. We focused on three SD-related organs that are relevant in the tilapia SD cascade: brain, testis, and pituitary. Interestingly, 17 out of the 39 genes showed overexpression in the brain; expression of 15 of these genes exceeded the median expression by over 3-fold. Thirteen genes were found relevant to SD following a literature survey. After removing 3 genes with no expression data, 4 out of the remaining 10 genes showed overexpression in at least one SD-associated organ (*Notch2*, *PIAS4*, *ZBTB7*, or *CELF5*).

Comparative mapping

Comparative analysis of the genes positioned within the SR detected high level of orthology between tilapia and six different species. Within <1.3 Mbp region of stickleback, Tetraodon, fugo, zebrafish, medaka, and human, 40, 39, 29, 29, 29, and 21 orthologous genes,

Table 4 SD-related data for annotated genes in the SR on scaffold 101 between 990,577 and 2,468,000 bp

	Gene or Symbol	Accession or Ensembl No.	Scaffold 101 (bp)		SD-related Publications	Gene Expression		
			Start	End		Brain	Testis	Pituitary
1	*PLIN3*	NP_001167399.1	1187253	1192169				
2	*ZFAND6*	XP_002199446.1	1194645	1198048			√√	
3	*FAM108C1*	XP_001342996.2	1205385	1210274				
4	*RGL1*	NP_991200.1	1216521	1224747		√√	√	
5	*GLT25D2*	ENSGACT00000016402	1245440	1265427		√√		
6	Novel protein	ENSGACT00000016408	1273512	1277395			ND	
7	*C1orf21*	ENSGACT00000016410	1273570	1287700		√√		
8	*EDEM3*	XP_688275.4	1305007	1311942				
9	*NPL*	NP_001133311.1	1314238	1318468				
10	*SEC22B*	ACM09163.1	1328257	1332317				√√
11	*NOTCH2*[a]	NP_001108566.1	1348777	1380740	Zhu *et al.* 2007		√√	
12	*SLC35A3*	ACN10890.1	1384970	1391950				
13	*FAM78B*	CAQ14615.1	1390770	1404535			ND	
14	*C19orf60*	NP_001158740.1	1406814	1410851		√√		
15	*CRLF1*	NP_001002650.1	1412045	1421020				
16	*TMEM59L*	ENSGACT00000016558	1422768	1431927		√√		
17	Novel protein	ENSGACT00000016563	1431927	1433134			ND	
18	*SSBP4*	NP_001018403.1	1502998	1511875		√√		
19	*FKBP8*	NP_001133417.1	1569993	1579098		√√		
20	*ELL*[a]	NP_956001.1	1603737	1613759	Zhou *et al.* 2009			
21	*DOT1*	CAP09616.1	1667805	1679590				
22	*Amh*	ABS58513.1	1688658	1695299			√√	
23	Novel protein	ENSGACT00000016737	1691664	1695317			ND	
24	*OAZ1*	NP_001134904.2	1696653	1701716		√√		
25	*dkey-3k20.4*	ENSGACT00000016747	1705254	1704891			ND	
26	*ORG*[b]	NP_001093540.1	1709490	1716883	Dai *et al.* 2009		ND	
27	*LINGO3*	ENSGACT00000016753	1769930	1771737				
28	Novel protein	ENSGACT00000016755	1783055	1787782			ND	
29	*ATP8B3*	XP_003201102.1	1808332	1824572				
30	*ONECUT3*	ENSGACT00000016776	1832236	1848746			ND	
31	*PIAS4*[a]	AAH57528.1	1876824	1885646	Hsieh *et al.* 2009	√√	√	
32	*MAP2K2*[a]	XP_002761634.1	1892758	1902715	Murakami *et al.* 2001			
33	*ZBTB7*[a]	CAK04316.1	1917084	1923570	Gailey *et al.* 2006	√√		
34	*TCF3*[a]	NP_001187227	1948472	1961119	Zhu *et al.* 2007			
35	*QCR10*[a]	ACQ58208.1	1968852	1973040			ND	
36	*MBD3b*	CAK10918.1	1973187	1979287	Kaji *et al.* 2006		ND	
37	*MYO5b*	CAK10917.1	1980557	1993470				
38	*UNC13A*	NP_001038630.1	1997106	2036658		√√		√
39	*HMG20b*[a]	NP_001018387.1	2039148	2044881	Sumoy *et al.* 2000			
40	*EEF2*	NP_956752.2	2047515	2063389				
41	*SNORD37*	ENSGACT00000029482	2050851	2051406			ND	
42	*Rxfp3*[b]	NP_001077348.1	2080930	2081856	Wilson *et al.* 2009			
43	*CREB3*	NP_001018509.1	2088906	2099759		√√		√
44	*CELF5*[a]	NP_001124260.1	2263338	2289589	Ladd *et al.* 2001	√√		
45	*RGMA*	NP_001133864.1	2299255	2306831		√√	√√	
46	*HSD11B1L*[b]	NP_001098261.1	2342628	2365654	Ozaki *et al.* 2006		ND	
47	*QIL1*	ACI69344.1	2365927	2368604		√	√	
48	*SPIN1*	XP_001339043.4	2370662	2372447		√		
49	*CFD*	ACI69308.1	2392555	2404506			√	
50	*BTBD8*[a]	EAW73104.1	2406880	2407219	Couderc *et al.* 2002			
51	*CYLIP1*	ACN58730.1	2459356	2488052		√√		

√ indicates 2–3× from median expression in BioGPS; √√ indicates >3× from median expression in BioGPS; ND, no data; SD, sex determination; SR, sex region.
[a] Gene related to SD in transcriptional processes.
[b] Gene associated with gonad development/function.

respectively, have been found in the same order. GO term enrichment analysis of these genes with DAVID software (Huang *et al.* 2009) yielded 4 genes, *Notch2*, *ELL*, *Amh*, and *TCF*, which are involved in biological processes of cell differentiation, cellular, and anatomical structure development based on zebrafish background of 8389 genes.

DISCUSSION

Different SD systems with remarkable variation have been observed in teleosts (Volff and Schartl 2002). Evidence supports that sex determination (SD) in tilapia is controlled by major genetic factors that may interact with minor genetic as well as environmental factors, thus implying that SD should be analyzed as a quantitative trait.

QTL for SD in *Oreochromis niloticus* were previously detected on LG1 and LG23 (Lee *et al.* 2003; Cnaani *et al.* 2004; Eshel *et al.* 2010). In the present study, interval mapping analysis using 33 markers on LG23 detected the QTL peak between two adjacent genetic markers: ARO172 and ARO177. However, the confidence interval was still rather large between 5 cM (Eshel *et al.* 2010; 156 individuals) and 30 cM in the current study. Thus, mapping QTL to confidence interval < 5 cM is not a viable option using genetic markers and segregating families of moderate size (Ron and Weller 2007). However, using physical mapping based on the unpublished tilapia genome sequence, all 26 markers in the QTL were physically mapped to three scaffolds on LG23. Furthermore, recombinations in two females were used to identify the boundaries of the SR between markers GM597 and ARO124 on a single scaffold (no. 101; Figure 2). This explains the lack of power of the interval mapping that is based on bootstrap analysis of a family of 90 individuals of which only two are informative. Figure 2 demonstrates the distinct contrast of genotypes for markers between genders in the specific sex region. The absence of recombination along a region of 12 genetic markers may reflect the moderate size of the family, but it also conforms to the theory that the evolution of sex chromosomes involves suppressed recombination between homologous chromosomes to maintain sex-related gene blocks (Bergero and Charlesworth 2009). The SR encompasses 1.5 Mbp harboring 51 annotated genes. Our assumption that the SR harbors sex-related or male-determining genes is strengthened by the conservation of this region in other teleost fish. Out of 51 genes that were positioned within the SR, 40 and 39 orthologous genes have been found within <1.3 Mbp region of stickleback and Tetraodon, respectively. Information from the literature indicates the putative role of 13 out of the 51 genes in SD: 10 genes in transcriptional processes related to SD and 3 in gonadal development and function (Table 4).

We examined expression of genes in the SD pathways at early developmental stages of tilapia. Previous studies on SD-related gene expression in tilapia focused on brain, PGS, and gonads (Ijiri *et al.* 2008; Poonlaphdecha *et al.* 2011). The results from our study on expression data of 10 candidate genes indicate that the onset of the SD cascade begins at 2 dpf at the gastrulation stage, based on overexpression of *Lhx9* and *ELAVL1* in females. *Lhx9* was found to be essential for mouse gonad formation (Birk *et al.* 2000). *ELAVL1* is a member of CELF proteins implicated in cell-specific and developmentally regulated alternative splicing (Ladd *et al.* 2001). Additional SD-related genes were *Sox9*, which is necessary and sufficient to cause testicular differentiation in mammals (Vidal *et al.* 2001). Likewise, *Foxl2* plays a role in ovarian sex differentiation and has been suggested to function as a repressor of the male pathway during ovarian development prophase (Ottolenghi *et al.* 2005). Significant differences in expression between genders for *Sox9a*, *Sox9b*, *Foxl2*, and *Sox14* genes were detected in later stages of embryonic development and may indicate their downstream role in the SD cascade. We detected higher expression of *SOX9* in females than in males at 7 dpf, in contrast to the results of Ijiri *et al.* (2008) of higher expression in male gonads at 37–70 days posthatching. *Foxl2* was also highly expressed in females at 8 dpf as was previously reported (Ijiri *et al.* 2008). Among 10 ten candidate genes, *Amh*, which is located in the center of the SR, showed the highest expression in male *vs.* female embryos. Our observation was supported by Poonlaphdecha *et al.* (2011) who reported on dimorphic expression of *Amh* between genders in adult gonads and brains as well as in embryo heads at 10 and 15 dpf. GO term enrichment analysis detected 4 genes, including *Amh*, that are involved in biological processes of cell differentiation, cellular development, and anatomical structure development. Genes playing a role in SD initiation with dimorphic expression between genders may be considered as candidate genes and should be further investigated.

To test the role of *Amh* and other candidate genes in SD of tilapia, targeted strategies could be considered, such as (i) mutant detection in candidate genes, as performed in zebrafish (Demarest *et al.* 2011); (ii) gene silencing using siRNA technology, as applied in the giant freshwater prawn (Ventura *et al.* 2009); and (iii) transgenesis using the *Tol2* system, which was demonstrated for Nile tilapia (Fujimura and Kocher 2011). Large-scale experiments might involve (i) genomic mutagenesis together with sex reversal, phenotypic mutant screening, and sequence analysis, as was applied in a medaka SD study (Otake *et al.* 2006); and (ii) a whole-transcriptome scan for gene expression at early embryonic development to identify the key regulators of SD. A complete computational approach was pursued to design a 44k features microarray (O. Eshel, unpublished data) based on the unpublished tilapia genome sequence annotation and EST libraries (Lee *et al.* 2010) for construction of the full tilapia gene list.

ACKNOWLEDGMENTS

We thank Berta Levavi-Sivan for providing us the ΔXX males developed in her laboratory. We thank the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program, Federica Di Palma, and Kerstin Lindblad-Toh for making the unpublished genome sequence data for *Oreochromis niloticus* available. This work is a contribution from the Institute of Animal Science, Agricultural Research Organization, Bet Dagan, Israel, No. 592/11. This research was supported by Grant IS-3995-07 from the United States–Israel Binational Agricultural Research and Development (BARD) Fund and by Grant No. 801/11 from the Israeli Science Foundation.

LITERATURE CITED

Albertson, R. C., J. T. Streelman, and T. D. Kocher, 2003 Directional selection has shaped the oral jaws of Lake Malawi cichlid fishes. Proc. Natl. Acad. Sci. USA 100: 5252–5257.

Baroiller, J. F., H. D'Cotta, and E. Saillant, 2009 Environmental effects on fish sex determination and differentiation. Sex Dev. 3: 118–135.

Bellott, D. W., H. Skaletsky, T. Pyntikova, E. R. Mardis, T. Graves,, *et al.* 2010 Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466: 612–616.

Bergero, R., and D. Charlesworth, 2009 The evolution of restricted recombination in sex chromosomes. Trends Ecol. Evol. 24: 94–102.

Birk, O. S., D. E. Casiano, C. A. Wassif, T. Cogliati, L. P. Zhao *et al.*, 2000 The LIM homeobox gene *Lhx9* is essential for mouse gonad formation. Nature 403: 909–913.

Bull, J. J., 1981 *Evolution of Sex Determining Mechanisms*. Benjamin-Cummings, California.

Charlesworth, D., and J. E. Mank, 2010 The birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics 186: 9–31.

Cnaani, A., and B. Levavi-Sivan, 2009 Sexual development in fish, practical applications for aquaculture. Sex Dev. 3: 164–175.

Cnaani, A., N. Zilberman, S. Tinman, G. Hulata, and M. Ron, 2004 Genome-scan analysis for quantitative trait loci in an F-2 tilapia hybrid. Mol. Genet. Genomics 272: 162–172.

Cnaani, A., B. Y. Lee, C. Ozouf-Costaz, C. Bonillo, F. Baroiller *et al.*, 2007 Mapping of *Sox2* and *Sox14* in tilapia (*Oreochromis* spp.). Sex Dev. 1: 207–210.

Cnaani, A., B. Y. Lee, N. Zilberman, C. Ozouf-Costaz, G. Hulata *et al.*, 2008 Genetics of sex determination in tilapiine species. Sex Dev. 2: 43–54.

Couderc, J. L., D. Godt, S. Zollman, J. Chen, M. Li *et al.*, 2002 The bric a brac locus consists of two paralogous genes encoding BTB/POZ domain proteins and acts as a homeotic and morphogenetic regulator of imaginal development in Drosophila. Development 129: 2419–2433.

Dai, L., W. Ma, J. Li, Y. Xu, W. Li *et al.*, 2009 Cloning and characterization of a novel oocyte-specific gene zorg in zebrafish. Theriogenology 71: 441–449.
Demarest, B. L., W. H. Horsley, E. E. Locke, K. Boucher, D. J. Grunwald *et al.*, 2011 Trans-centromere fffects on meiotic recombination in the zebrafish. Genetics 187: 333–336.
Devlin, R. H., and Y. Nagahama, 2002 Sex determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. Aquaculture 208: 191–364.
Ellegren, H., 2011 Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nat. Rev. Genet. 12: 157–166.
Eshel, O., A. Shirak, J. I. Weller, T. Slossman, G. Hulata *et al.*, 2010 Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). Anim. Genet. 42: 222–224.
Fujimura, K., and T. D. K. Kocher, 2011 Tol2-mediated transgenesis in tilapia (*Oreochromis niloticus*). Aquaculture 319: 342–346.
Fujimura, K., and N. Okada, 2007 Development of the embryo, larva and early juvenile of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae). Developmental staging system. Dev. Growth Differ. 49: 301–324.
Gailey, D. A., J. C. Billeter, J. H. Liu, F. Bauzon, J. B. Allendorfer *et al.*, 2006 Functional conservation of the fruitless male sex-determination gene across 250 Myr of insect evolution. Mol. Biol. Evol. 23: 633–643.
Hsieh, H. T., C. H. Wang, M. L. Wu, F. M. Yang, Y. C. Tai *et al.*, 2009 PIASy inhibits LRH-1-dependent CYP11A1 expression by competing for SRC-1 binding. Biochem. J. 419: 201–209.
Huang, D. W., B. T. Sherman, and R. A. Lempicki, 2009 Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat. Protoc. 4: 44–57.
Ijiri, S., H. Kaneko, T. Kobayashi, D. S. Wang, F. Sakai *et al.*, 2008 Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. Biol. Reprod. 78: 333–341.
Kaji, K., I. M. Caballero, R. MacLeod, J. Nichols, V. A. Wilson *et al.*, 2006 The NuRD component Mbd3 is required for pluripotency of embryonic stem cells. Nat. Cell Biol. 8: 285–292.
Kasahara, M., K. Naruse, S. Sasaki, Y. Nakatani, W. Qu *et al.*, 2007 The medaka draft genome and insights into vertebrate genome evolution. Nature 447: 714–719.
Knott, S. A., J. M. Elsen, and C. S. Haley, 1996 Methods for multiple-marker mapping of quantitative trait loci in half-sib populations. Theor. Appl. Genet. 93: 71–80.
Ladd, A. N., N. Charlet-B, and T. A. Cooper, 2001 The CELF family of RNA binding proteins is implicated in cell-specific and developmentally regulated alternative splicing. Mol. Cell. Biol. 21: 1285–1296.
Lee, B. Y., D. J. Penman, and T. D. Kocher, 2003 Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. Anim. Genet. 34: 379–383.
Lee, B. Y., G. Hulata, and T. D. Kocher, 2004 Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). Heredity 92: 543–549.
Lee, B. Y., W. J. Lee, J. T. Streelman, K. L. Carleton, A. E. Howe *et al.*, 2005 A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). Genetics 170: 237–244.
Lee, B. Y., A. E. Howe, M. A. Conte, H. D'Cotta, E. Pepey *et al.*, 2010 An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. BMC Genomics 11: 278.
Mair, G. C., J. S. Abucay, D. O. F. Skibinski, T. A. Abella, and J. A. Beardmore, 1997 Genetic manipulation of sex ratio for the large-scale production of all-male tilapia, *Oreochromis niloticus*. Can. J. Fish. Aquat. Sci. 54: 396–404.
Mank, J. E., D. E. L. Promislow, and J. C. Avise, 2006 Evolution of alternative sex-determining mechanisms in teleost fishes. Biol. J. Linn. Soc. Lond. 87: 83–93.
Otake, H., A. Shinomiya, M. Matsuda, S. Hamaguchi, and M. Sakaizumi, 2006 Wild-derived XY sex-reversal mutants in the medaka, *Oryzias latipes*. Genetics 173: 2083–2090.
Ottolenghi, C., S. Omari, J. E. Garcia-Ortiz, M. Uda, L. Crisponi *et al.*, 2005 *Foxl2* is required for commitment to ovary differentiation. Hum. Mol. Genet. 14: 2053–2062.
Ozaki, Y., M. Higuchi, C. Miura, S. Yamaguchi, Y. Tozawa *et al.*, 2006 Roles of 11 beta-hydroxysteroid dehydrogenase in fish spermatogenesis. Endocrinology 147: 5139–5146.
Palti, Y., A. Shirak, A. Cnaani, G. Hulata, R. R. Avtalion *et al.*, 2002 Detection of genes with deleterious alleles in an inbred line of tilapia (*Oreochromis aureus*). Aquaculture 206: 151–164.
Poonlaphdecha, S., E. Pepey, S. H. Huang, M. Canonne, L. Soler *et al.*, 2011 Elevated amh gene expression in the brain of male tilapia (*Oreochromis niloticus*) during testis differentiation. Sex Dev. 5: 33–47.
Ron, M., and J. I. Weller, 2007 From QTL to QTN identification in livestock–winning by points rather than knock-out: a review. Anim. Genet. 38: 429–439.
Rougeot, C., C. Prignon, C. V. N. Kengne, and C. Melard, 2008 Effect of high temperature during embryogenesis on the sex differentiation process in the Nile tilapia, *Oreochromis niloticus*. Aquaculture 276: 205–208.
Schartl, M., 2004 Sex chromosome evolution in non-mammalian vertebrates. Curr. Opin. Genet. Dev. 14: 634–641.
Shirak, A., Y. Palti, A. Cnaani, A. Korol, G. Hulata *et al.*, 2002 Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). J. Hered. 93: 270–276.
Shirak, A., E. Seroussi, A. Cnaani, A. E. Howe, R. Domokhovsky *et al.*, 2006 Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. Genetics 174: 1573–1581.
Spelman, R. J., W. Coppieters, L. Karim, J. A. M. vanArendonk, and H. Bovenhuis, 1996 Quantitative trait loci analysis for five milk production traits on chromosome six in the Dutch Holstein-Friesian population. Genetics 144: 1799–1807.
Su, A. I., T. Wiltshire, S. Batalov, H. Lapp, K. A. Ching *et al.*, 2004 A gene atlas of the mouse and human protein-encoding transcriptomes. Proc. Natl. Acad. Sci. USA 101: 6062–6067.
Sumoy, L., L. Carim, M. Escarceller, M. Nadal, M. Gratacos *et al.*, 2000 HMG20A and HMG20B map to human chromosomes 15q24 and 19p13.3 and constitute a distinct class of HMG-box genes with ubiquitous expression. Cytogenet. Cell Genet. 88: 62–67.
Ventura, T., R. Manor, E. D. Aflalo, S. Weil, S. Raviv *et al.*, 2009 Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. Endocrinology 150: 1278–1286.
Vidal, V. P. I., M. C. Chaboissier, D. G. de Rooij, and A. Schedl, 2001 *Sox9* induces testis development in XX transgenic mice. Nat. Genet. 28: 216–217.
Volff, J. N., and M. Schartl, 2002 Sex determination and sex chromosome evolution in the medaka, *Oryzias latipes*, and the platyfish, *Xiphophorus maculatus*. Cytogenet. Genome Res. 99: 170–177.
Wilson, B. C., D. Burnett, R. Rappaport, L. J. Parry, and E. K. Fletcher, 2009 Relaxin-3 and RXFP3 expression, and steroidogenic actions in the ovary of teleost fish. Comp. Biochem. Physiol. A Mol. Integr. Physiol. 153: 69–74.
Wu, C., C. Orozco, J. Boyer, M. Leglise, J. Goodale *et al.*, 2009 BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. Genome Biol. 10: R130.
Zilberman, N., S. Reikhav, G. Hulata, and M. Ron, 2006 High-throughput genomic DNA extraction protocol from tilapia's fin tissue. Aquaculture 255: 597–599.
Zhou, J. G., X. Feng, B. Ban, J. X. Liu, Z. Wang *et al.*, 2009 Elongation factor ELL (eleven-nineteen lysine-rich leukemia) acts as a transcription factor for direct thrombospondin-1 regulation. J. Biol. Chem. 284: 19142–19152.
Zhu, X. Y., A. S. Gleiberman, and M. G. Rosenfeld, 2007 Molecular physiology of pituitary development: signaling and transcriptional networks. Physiol. Rev. 87: 933–963.

Communicating editor: D.-J. De Koning

Part C - Functional mapping[1]

Identification of Male-specific *amh* Duplication, Sexually Differentially Expressed Genes and microRNAs at Early Embryonic Development of Nile tilapia (*Oreochromis niloticus*)

Orly Eshel[1,2], Andrey Shirak[2], Lior Dor[2], Mark Band[3], Tatyana Zak[4], Michal Markovich-Gordon[5], Vered Chalifa-Caspi[5], Esther Feldmesser[6], Joel I. Weller[2], Eyal Seroussi[2], Gideon Hulata[2], Micha Ron[2,*]

[1]Robert H. Smith Faculty of Agriculture, Food and Environment, Hebrew University of Jerusalem, Rehovot 76100, Israel.
[2]Institute of Animal Science, ARO, The Volcani Center, 50250, Israel.
[3]W.M. Keck Center for Comparative and Functional Genomics, University of Illinois at Urbana Champaign, Urbana, IL, USA.
[4]Fish & Aquaculture Research Station, Dor, Hof HaCarmel, 30820, Israel.
[5]National Institute of Biotechnology in the Negev, Ben-Gurion University of the Negev, Beer-Sheva, Israel.
[6]The Nancy and Stephen Grand National Center for Personalized Medicine, Weizmann Institute of Science, Rehovot, Israel.

Orly Eshel: orly.eshel@mail.huji.ac.il
Andrey Shirak: shiraka@volcani.agri.gov.il
Lior Dor: lior.dor@mail.huji.ac.il
Mark Band: markband@illinois.edu
Tatiana Zak: Taniyaz@moag.gov.il
Michal Markovich-Gordon: gordonmi@exchange.bgu.ac.il
Vered Chalifa-Caspi: veredcc@bgu.ac.il
Esther Feldmesser: ester.feldmesser@weizmann.ac.il
Joel I. Weller: Joel.weller@mail.huji.ac.il
Eyal Seroussi: eyal.seroussi@mail.huji.ac.il
Gideon Hulata: vlaqua@volcani.agri.gov.il
Corresponding author, Micha Ron: micha.ron@mail.huji.ac.il

Abstract

Background: The probable influence of genes and the environment on sex determination in Nile tilapia suggests that it should be regarded as a complex trait. Detection of sex determination genes in tilapia has both scientific and commercial importance. The main objective was to detect genes and microRNAs that were differentially expressed by gender in early embryonic development.

Results: Artificial fertilization of *Oreochromis niloticus* XX females with either sex-reversed ΔXX males or genetically-modified YY 'supermales' resulted in all-female and all-

[1]Submitted to BMC Genomics; accepted with minor revision

male embryos, respectively. RNA of pools of all-female and all-male embryos at 2, 5 and 9 dpf were used as template for a custom Agilent eArray hybridization and next generation sequencing. Fifty-nine genes differentially expressed between genders were identified by a false discovery rate of $p<0.05$. The most overexpressed genes were *amh* and *tspan8* in males, and *cr/20ß-hsd*, *gpa33*, *rtn4ipl* and *zp3* in females ($p<1\times10^{-9}$). Validation of gene expression using genomic qPCR indicated copy number variation in *tspan8*, *gpa33*, *cr/20ß-hsd* and *amh*. Sequencing of *amh* identified a male-specific duplication of this gene, denoted *amhy*, differing from the sequence of *amh* by a 233 bp deletion on exonVII, hence lacking the capability to encode the protein motif that binds to the transforming growth factor beta receptor (TGF-β domain). *Amh* and *amhy* segregated in the mapping family in full concordance with SD-linked marker on LG23 signifying the QTL for SD. We discovered 831 microRNAs in tilapia embryos of which nine had sexually dimorphic expression patterns by a false discovery rate of $p<0.05$. An up-regulated microRNA in males, pma-mir-4585, was characterized with all six predicted target genes including *cr/20ß-hsd*, down-regulated in males.

Conclusions: This study reports the first discovery of sexually differentially expressed genes and microRNAs at a very early stage of tilapia embryonic development, i.e. from 2 dpf. Genes with sexually differential expression patterns are enriched for copy number variation and apoptosis. A novel male-specific duplication of *amh*, denoted *amhy*, lacking the TGF-β domain was identified and mapped to the QTL region on LG23 for SD, thus indicating its potential role in SD.

Keywords: sex-determination, *amh*, *amhy*, microRNA, CNV, *cr/20ß-hsd*, tilapia

Background

There are more than 24,000 species of fish [1]. Research on fish SD has provided important insight into the plasticity of the sex-determination process in vertebrates since the biology and ecology of fish is particularly diverse and provides unique examples of sex-determination mechanisms, yet they possess many of the same processes and pathways that are used in other vertebrate systems. Sex and sex ratio have been attributed to a dominant gene (*SRY* in human), gene dosage (*Drosophila*), environmental influence (Alligator), or by the 'threshold dichotomy' theory, that applies to a trait with contrasting phenotypes originating from multiple genes with quantitative effects [2-5]. Tilapia SD has been well

studied for its potential to produce all-male progeny with enhanced growth rate due to lack of reproductive interactions in commercial ponds. However, dimorphic differences between male and female karyotpes have not been displayed [6]. A variety of evidence suggests that sex determination in tilapia is a complex trait governed by the interactions between a genetic determination and the influence of temperature [7]. The hypothesized dual sex chromosome system for tilapia species, XX-XY system for *O. mossambicus* and *O. niloticus*, and WZ-ZZ system for *O. aureus* and *O. urolepis hornorum* was adopted by Hickling [8]. The primary support for these hypotheses was obtained from analysis of sex-ratio of progeny of: i. inter-specific crosses [9]; ii. intra-specific crosses using sex-reversed individuals [10]; and iii. chromosome set manipulations through gynogenesis [11] and androgenesis [12].

The differences in the SD mechanism among closely related tilapia species and the probable influence of sex determining genes and the environment, suggest that SD should be analyzed using a markers-based QTL approach [7,13]. However, *O. niloticus* and *O. aureus* have different sex chromosome systems and their ability to mate and produce fertile hybrids further complicates the elucidation of the SD system. Mapping QTL for SD was based on a second-generation genetic linkage map of tilapia [14]. Studies in *O. aureus, O. mossambicus*, *O. niloticus* and F_2 family derived from *O. aureus* x *O. niloticus* cross identified QTL for SD on LG 1, 3 and 23 [15-20]. The region on LG23 affecting sex was further fine mapped using a segregating family of Nile tilapia to 1.47 Mbp harboring 51 genes including *amh* [21]. Differential expression of *amh* between genders was reported in brain and gonads from 10 days post fertilization (dpf) embryos in Nile tilapia [22]. In zebrafish *amh* has independent functions in inhibiting both steroidogenesis and spermatogenesis [23]. The complexity of SD and the limitations of QTL mapping and the candidate gene approach [24] complicate the identification of the causative genes for SD. Thus, transcriptome-wide gene expression by gender may be used for the identification of genes that are involved in SD and sex differentiation.

microRNA (miRNA) are small noncoding RNAs, about 21 nucleotides in length. Many are conserved, and may regulate up to 30% of gene expression by base-pairing to partially complementary mRNAs [25]. Recently, Huang et al. [26] published 184 miRNAs in skeletal muscle of Nile tilapia. Yan et al. [27] identified 25 conserved miRNAs in tilapia skeletal muscle using small RNA cloning. By examining the expression of nearly 250 of the most abundant rodent miRNAs, Bale and Morgan [28] identified a robust sex-specific pattern of miRNA expression in the neonatal brain. Study on mouse characterized 55 miRNA signatures in testis and ovary [29] and illustrated their importance for the proliferation of

Table

PGCs and spermatogonia [30]. Additional studies in chicken and zebrafish identified sex-specific pattern of miRNA expression in brain, embryo and gonads [31,32]. These findings suggest that miRNAs may play a significant role in development and more specifically in SD.

The critical period of sensitivity for elevated temperature [33] or hormonal treatment [34] to induce sex reversal of Nile tilapia was determined from fertilization to 21 days post hatching. Ijiri et al. [35] detected differentially expressed genes in XX and XY bi-potential gonads during the period of 9–10 dpf. Rougeot et al. [36] applied temperature treatment on presumable all-female population embryos until hatching (2-3 dpf) and showed ~20% phenotypic sex reversal of females to males. In addition, the findings of sex-specific mortality closely after hatching indicates that the initiation and regulation of SD pathways begin during the first few days of embryonic development, i.e. <3 dpf [15]. Preliminary analysis of candidate genes for SD at 2 to 9 dpf confirmed their functionality at early embryonic development [21]. Thus, the objective of this study was to conduct a transcriptome-wide search in Nile tilapia at early embryonic development for genes and miRNA of the SD mechanism and sex differentiation.

Methods

Animals and tissue collection

Breeding of *Oreochromis niloticus* (Swansea stock) families used in this study was performed at the aquaculture research station Dor, Israel. To obtain all-female (XX) and all-male (XY) progeny, eggs collected from six *O. niloticus* females were artificially fertilized with milt stripped from either two hormonally sex-reversed males (ΔXX, Nile tilapia, Manzala strain) or three genetically-modified 'supermales' (YY, Nile tilapia, Swansea strain, Fishgen Ltd) thus creating all-female and all-male progeny, respectively [37]. For each full-sib group, a pool of 15-30 embryos were collected at 2, 5 and 9 dpf, immediately placed in RNAlater reagent (Qiagen, USA) and then stored at −20°C until RNA extraction. The remaining fish in the group were grown until the age of three months and the sex of at least 60 individuals was determined by microscopic analysis (X100 magnification) of gonadal squash. Groups with less than 95% of individuals having the same sex were not included in the experiment. At 75 dpf, five males and females from each full-sib group were sacrificed for collection of brain, gonads and liver. The experimental protocol was approved by the Animal Care Committee of ARO.

DNA, RNA extraction and cDNA synthesis

DNA was extracted from fin samples using the MasterPure™ DNA Purification Kit (Epicentre® Biotechnologies, WI, USA) following the manufacturer's recommended protocol. Total RNA was extracted from a pool of 5-15 deyolked embryos (mirVana™ miRNA Isolation Kit, Ambion). Synthesis of cDNA was done with SuperScript II (Invitrogen, USA) according to the manufacturer's instructions. The quantity and quality of the RNA samples were verified using a NanoDrop ND-1000 Spectrophotometer (NanoDrop Technologies, Wilmington, DE) and an Agilent 2100 Bioanalyzer for RNA (Agilent Technologies, Palo Alto, CA).

Agilent microarray design

In the absence of an expression array in tilapia we searched three bioinformatic resources of tilapia, i.e. *O. niloticus*; Broad Institute assembly (Orenil1.1; accession no. PRJNA59571), EST libraries [38] and candidate genes for SD to establish a 35,156, 5,561 and 696 different probes, respectively that represents the transcriptome (Figure 1). Custom gene expression array was designed using eArray (Agilent Technologies, Santa Clara, CA) and 60-mer probe synthesis on a 4 × 44 k format. Our target sequences for generation of probes were based on the tilapia genome and the EST libraries. Genes were annotated by the Maker pipeline that is based on *ab-initio* gene predictions and EST evidence, and *ab-initio* SPAN gene predictions based on a full genome training set. Additional genes were identified based on assembly of tilapia EST libraries (using MIRA [39]) followed by BLASTX to fish ORF [40]. Additionally, a set of 103 positive and negative probes were designed from known genes, and were represented 6 to 7 times each on the array. Negative controls were probes representing three plant genes, whereas positive controls were probes of genes that were known to participate in sex determination and/or differentiation pathways in various species (*cyp19*, *sox9*, *amh*, *elavl1*, *dmrt1*, *foxl2*, *lhx9*, *sox14*, *msp* and *gnrh2*) and genes spanning from 1,050 to 2,488 Kbp on scaffold 102 of tilapia that were previously suggested as positional candidate genes for SD [21,41]. The resulting eArray of 43,803 probes was used for hybridization with cDNA of 56 biological samples of predetermined gender at 2, 5 and 9 dpf of embryonic development.

Microarray hybridization and data analysis

Two hundred ng of total RNA was labeled using the Agilent 2-color low input quickamp labeling kit according to the manufacturer's protocols (Agilent Technology, Santa

Clara, CA). Labeled samples were hybridized to a custom designed tilapia 4 x 44K eArray containing 43,803 probes and scanned on an Axon 4000B microarray scanner (Molecular Devices, Sunnyvale, CA) at 5 μm resolution. A total of 56 biological samples at three embryonic developmental stages of 2, 5 and 9 dpf were used as template for cDNA synthesis and array hybridization. Spot finding and background correction of signal intensities were carried out using GenePix 6.1 software (Molecular Devices, Sunnyvale, CA). The microarray expression data was normalized with the Bioconductor 2.8 LIMMA package [42] using loess and aquantile normalization with a single channel analysis design. The replicated probes were represented by their median expression for analysis. The microarray data were deposited in NCBI's Gene Expression Omnibus [43] under accession No. GSE50974.

Statistical analysis

The normalized data of each of 43,210 different probes recorded on 56 samples were log transformed and PCA was performed by the Partek software [44] with normalized eigenvector scaling and correlation dispersion matrix. In addition each probe was analyzed by the General Linear Model (PROC GLM) procedure of SAS. The effects included in the model were slide, array, dye, gender (male or female), sire nested within gender, dpf and dam. Least square means were computed for the effects of gender. Gender was given a value of 0 for males and 1 for females. Probabilities of the differences between the "male" and "female" effects were computed based on the least square means standard errors for each effect. The probability values were then sorted from lowest to highest, and the false discovery rate (FDR) was computed for each probe, based on the ratio of expected to observed numbers of probes for each probability value. Bonferroni probabilities taking into account multiple testing were also computed as 1- the Poisson probability to obtain zero "significant" probes for each expectation of the number of probes for each nominal probability value. Pearson correlations were computed among the 59 significant probes for gender effects at each time point for the three pair-wise combinations of dpf (2, 5 and 9).

Functional annotation clustering

The tilapia genome is not well annotated compared to the human genome. Thus, we used the human orthologs for gene ontology analyses. The corresponding human Gene IDs were identified using NCBI BLAST. Thirty-nine out of 59 genes had identified human

orthologs (Table S1). The DAVID classification system [45] was performed to assess the probability of over representation of genes within the list of human orthologs of certain pathways, biological processes and molecular functions using medium to highest classification stringencies and the default of Bonferroni correction for multiple testing.

Validation of microarray results using quantitative real-time PCR

Validation of probes was done by qPCR using RNA of three to four pools of monosex embryos, each comprising samples of 28 females or 30 males at eight daily time points from 2 to 9 dpf. Primer design were based on sequence of ~100 bp flanking each of eight probes that were submitted to ''Primer3plus''program [46]. For a given probe, one of the primers in each pair was identical to the original sequence from the microarray experiment targeting the probe's location. The fragment's sequence was used to BLAST search the tilapia genome to confirm its position in the genome (Table S1).

The qPCR analysis was performed in triplicates using the Fast SYBR® Green Master Mix kit (Thermo Fisher Scientific, UK) according to the instructions of the manufacturer in a 17-µl reaction volume, which included 2 µl of DNA (30 ng/µl), 1 µl of each primer (10 pmol/µl), 4.5 µl of ultra pure water, and 8.5 µl of Absolute Blue SYBER Green ROX Mix. The qPCR reaction was performed in the following conditions: 20 seconds at 95°C for enzyme activation followed by 40 cycles of 3 sec at 95°C, 30 sec at 60°C using StepOnePlus™ Real-Time PCR System. Amplification was followed by melting-curve analysis to confirm specificity of products. A standard curve was generated for each gene using serial dilutions of the specific PCR product, for the absolute quantification method. The threshold cycle number (Ct) for each tested probe was used to quantify its relative abundance. The StepOne Software v2.2.1 (ABI) was used for the calculation of the relative quantities using Glyceraldehyde-3-phosphate dehydrogenase (*gapdh*) for normalization. The relative amount of the target RNA, designated as the input amount (IA) was determined by comparison with the corresponding standard curve for each sample (User Bulletin #2 ABI PRISM7700 Sequence Detection System, Applied Biosystems). The IA values were calculated as follows: $IA = [10^{((Ct - intercept)/slope)}]$, where Ct is the cycle threshold for unknown sample. The female to male ratio of expression was computed based on the mean IA of eight daily samples of male and female embryos from 2 to 9 dpf. Pearson correlations between the female to male ratio of expression from the microarray experiment and that from qPCR was performed for eight genes (*fcgrt*, *rtn4ip1*, *CUST_26098*, *zp3*, *gpa33*, *tspan8*, *cr/20ß-hsd* and *amh*) using Excel. *cr/20ß-hsd* and *amh* were further characterized for their expression by

qPCR in brain, liver and gonads of 75 dpf female and male fish. These genes have shown expression in brain and gonads in previous studies, and liver is used as negative control.

Copy number variation using quantitative real-time PCR

Determination of the relative copy number of eight probes was conducted using qPCR analysis based on genomic DNA (gDNA) template. Preliminary work was conducted on randomly selected 13 and 14 male and female gDNA samples, respectively. Gene copy number was normalized to the reference gene, Ribonuclease P protein subunit p30 (*rpp30*) that was used as a reference gene in human CNV studies [47]. Additionally, five offspring of crosses between two dams and five sires were analyzed for relative copy number.

***amh* sequencing and linkage mapping**

Full length *amh* gene and 2,000 bp flanking it was amplified in ~1,000 bp fragments with primer design based on the Ensembl sequence scaffold GL831234.1:1,686,017-1,697,999 (Table 4). PCR templates included DNA of XX female, XY and YY males. PCR reaction protocol was as presented previously except the use of the high-fidelity BIO-X-ACT Long DNA polymerase (Bioline, London, UK). PCR products were separated on agarose gels and stained with ethidium bromide. The DNA fragments were visualized with UV light and excised from the gel. DNA fragments were purified with the DNA gel extraction kit (Millipore, Bedford, MA) and then sequenced on 3730 DNA analyzer (Applied Biosystem, USA). Sequence trace files were assembled and analyzed with the GAP4 package [48]. Linkage mapping for SD was performed by genotyping the *O. niloticus* mapping family for *amh* and *amhy* (Table 1) and microsatellite *UNH898* [21].

Small RNA sequencing

Six small RNA libraries were prepared for 'super' pools of full-sib embryos of males and females at 2, 5 and 9 dpf, that were used for the microarray experiment, with Illumina's 'TruSeq' Small RNA Sample Prep kit. The libraries were size selected to 18-33 nucleotides fragments, quantitated by qPCR and divided into two samples that were sequenced on separate lanes for 41 cycles on a HiSeq2000 using a TruSeq SBS sequencing kit version 3. The sequences were analyzed with Casava1.8 (pipeline 1.9) yielding between 21 to 32 million sequences per sample.

Identification of miRNAs

The sequence reads were mapped to the tilapia genome and analyzed by miRDeep 2.0.0.5- mapper script, an algorithm based on the miRNA biogenesis model [49]. It aligns sequencing reads to potential hairpin structures in a manner consistent with Dicer processing, and assigns log-odds scores to estimate the probability that hairpins are true miRNA precursors. The output of this analysis is a scored list of miRNAs that passed the stringent score cut-off of four, which reflects a signal-to-noise ratio greater than 12. Expression levels were normalized to the size of each of the six libraries (per millions). The miRNAs were searched against and submitted to miRbase for miRNA gene name assignment. Novel miRNAs in tilapia were assigned "oni-mir number". miRNA data was deposited in miRbase.

Differential expression of miRNAs between genders

We used the miRDeep 2.0.0.5 quantifier script, a module that maps the deep sequencing reads falling into an interval of two nucleotides upstream and five nucleotides downstream of the mature/star sequences of the predefined miRNA precursors, to estimate the expression of the corresponding miRNAs in each sample. Read counts of mature and star sequences of $\geq$ four read counts were log transformed and the deviations between genders were analyzed in each of 2, 5 and 9 dpf, separately. The FDR was calculated assuming a normal distribution, i.e. comparing the realized number of deviations $\geq$ 4 standard deviations to those expected by random at $p=3.17\times10^{-5}$. Thus, deviations between genders $\geq$ 4 standard deviations corresponding to a FDR of 1.5% were considered statistically significant.

Detection of gene hosts for differentially expressed miRNAs

Gene hosts were searched for the conserved up-regulated mir-21 and mir-218 in males, using BLASTN with their precursor sequences against the sequence data of vertebrate species (Ensembl). ESTs from vertebrates were assembled in the vicinity of the miRNAs to identify their gene hosts. Tissue specific pattern of mRNA expression in zebrafish was analyzed for the identified gene hosts using BioGPS [50].

miRNA target prediction analysis

Nine miRNAs differentially expressed by gender were explored for potential gene targets represented by their mature sequences (3p/5p) in 59 differentially expressed genes between genders (Table 2). The 3`UTR sequences of the 59 genes were downloaded from

Biomart in Ensembl database. When the 3`UTR was not available, 2,482 bp downstream was used. This length is the third quantile of the known tilapia 3`UTR. Prediction of potential gene targets for miRNA was preformed with RNAhybrid and miRanda software [51,52]. Forty four gene targets that were predicted by both algorithms were considered potential miRNA gene targets.

Results

Identification of genes and miRNA affecting SD were based on comparative analysis between genders using the same biological samples, i.e. RNA of tilapia embryos of predetermined gender, from 2 to 9 dpf.

Analysis of differentially expressed genes between genders and CNV

Principal component analysis (PCA) for the microarray data showed that three factors explained 40, 20 and 9 % of the total variance. Individual samples of gender by dpf, relative to the three factors are plotted in Figure S1. There was a clear distinction between males and females except for two female samples at 2 and 9 dpf which were close to the male cluster. The effect of gender was highly correlated with the first factor, while the effect of dpf was correlated with the second factor. Correlations with the third factor were low for all effects considered.

The FDR as a function of nominal significance values and number of significant probes is plotted in Figure S2. There were 59 genes differentially expressed between genders significant with an FDR of 0.05, which corresponded to nominal probability of 10^{-5}. Fold change values between genders ranged from 1.2 to 4.2. The gene names, their annotations, least squares means by gender as fold change between genders (FC) and potential relevance to SD and apoptosis/immune response are presented in Table S1. Hierarchical clustering of all differentially expressed genes shows that only one third of them were highly expressed in males (Figure 2). The expression profiles of these genes were similar across the three time points of 2, 5 and 9 dpf, with correlations exceeding 0.79. Nevertheless, there is a tendency of earlier differentially expressed genes in females (2 dpf) than in males (5 dpf). For 39 of the genes, orthologs in human were found with full annotations. Functional annotation clustering indicated a significant enriched cluster of the immune response containing four genes e.g., *psmb8*, *fcgrt*, *gas7a7* and *zp3* (DAVID enrichment score 1.56; $p<0.01$). Ten

genes were involved in pathways known to affect SD, and more than half of the genes (22) were involved in apoptosis (Table S1).

The genes with the most significant sexually dimorphic expression patterns were: *carbonyl reductase-like 20β-hydroxysteroid dehydrogenase* (*cr/20β-hsd*; Fold change (FC) 3.5; $p=1\times10^{-17}$), *reticulon-4-interacting protein 1 homolog (rtn4ip1;* FC 2.5; $p=1\times10^{-17}$*), Tetraspanin-8* (*tspan8*; FC 0.2; $p=1\times10^{-16}$), *inositol monophosphatase 1* (*impa1*; FC 2.8; $p=1\times10^{-14}$), *zinc-finger bed domain- containing 3* (*zbed3*; FC 0.6; $p=1\times10^{-12}$) and *anti-müllerian hormone* (*amh*; FC 0.5; $p=1\times10^{-9}$) (Table S1). *cr/20β-hsd*, *rtn4ip1* and *impa1* were highly expressed in females, and *tspan8, zbed3* and *amh* were overexpressed in males. The daily expression by qPCR of *amh* and *cr/20β-hsd* in 2 to 9 dpf embryos is presented in Figure 3A demonstrating the overexpression in males and females, respectively. The expression of *amh* and *cr/20β-hsd* by qPCR in brain, liver and gonads of tilapia at 75 dpf is presented in Figure 3B. *cr/20β-hsd* and *amh* genes' overexpression were validated in the respective gonads, i.e. ovary and testis, while *amh* was also highly expressed in male brain.

There were 23 probes with experimental-wise Bonferroni probabilities < 0.05. Eight out of the 23 genes were sampled for validation by qPCR. Expression data for microarray and qPCR are given in Table 2. A correlation of 0.8 for mean expression ratio between genders was obtained between the microarray and qPCR, indicating a high rate of validation. Nevertheless, melting curve analysis showed multiple amplified fragments that were specific to one of the genders for a few of the genes, indicating potential copy number variation. Further analysis by qPCR based on gDNA of males and females for four of the eight genes, *amh*, *cr/20β-hsd, tspan8* and *gpa33,* demonstrated significant copy number differences between genders. Figure 4 shows that *tspan8* had more genomic copies in males while *cr/20β-hsd* and *gpa33* had more genomic copies in females. All four genes showed direct correlation between number of copies and expression level.

Identification of male specific *amh* duplication

As a candidate gene for SD, *amh* was represented by five different probes on the eArray. Analysis of the microarray data showed differential expression between genders only for one of the probes that was located upstream to exon VII. This probe was highly expressed in males. Thus, we amplified and sequenced the full length *amh* gene using four PCR amplicons of about 1,000 bp each (Figure 5A, primers on Table 1). Length of amplified products was similar for both genders for the first three PCR amplicons covering exon I to VI. In the fourth amplicon we identified in both genders a fragment of 1,048 bp

containing exon VII and possibly an additional somewhat smaller fragment attached to it. However, an additional fragment of 815 bp was found only in males and was termed *amhy*.

Sequencing of the two types of fragments revealed nearly identical sequences to *amh* exon VII with a deletion of 233 bp in *amhy* (Figure 5B). We sequenced the *amh* and *amhy* exon VII for XX, XY and YY unrelated individuals. The purple arrow indicates nucleotides T and C at position 1,274 of the full length *amh* in XX and XY, respectively, with both alleles present in YY. The green arrow indicates an A>G substitution in nucleotide 1,275 of *amhy* between XY and YY individuals. In *amhy*, nucleotide G was found at position 1,403 as opposed to T in *amh* in both XY and YY individuals as pointed by red arrows, with lack of alignment thereafter, indicating the start of the deletion which corresponds to transforming growth factor beta (TGF-β) binding domain (Figure 6A). Capability of translation of exon VII in *amhy*, as compared to *amh*, indicates a potential deletion of 86 amino acids and addition of 21 amino acids, due to a reading frame shift and disruption of a stop codon (Figure 6B). The potentially translated protein based on the partial *amhy* sequence exhibits at its end 12 of the 21 predicted additional amino acids.

PCR amplification with internal primers designed based on *amh* exon VII, on cDNA of male and female 3 and 6 dpf embryos, brain, liver and gonads, showed the *amh* fragment of 442 bp in all samples except liver, whereas the *amhy* smaller fragment of 209 bp was detected in male embryos, brain and testis but not in female embryos, brain and ovary (Figure 5C). Interestingly, the intermediate fragment that may be observed in Figure 5A (from DNA) do not appear in Figure 5C (from RNA). It may be due to an artifact or another duplication.

Linkage mapping analysis showed full concordance between UNH898 on LG23, which is the closest marker to the QTL for SD [16, 21], *amh* [41] and *amhy*. Both male-associated-allele of UNH898 and *amhy* fragment were present in all 61 males and absent in all 29 females, thus indicating that *amhy* is localized to LG23 at the SD region.

Analysis of differentially expressed miRNAs between genders

The 171.2 million reads from the small RNA sequencing experiment were uploaded to miRDeep2 software which processes reads and using the Mapper script maps them to the reference genome for miRNA detection based on their biogenesis model. We discovered 578 miRNA precursors that passed the stringent score cut-off of four, which reflects a signal-to-noise ratio greater than 12 in tilapia and submitted them to miRBase.

We then ran the quantifier script to determine and normalize the number of reads of predefined miRNA precursors, indicating the expression of the corresponding miRNAs in each sample. Read counts by gender of 704, 668 and 636 mature and star sequences of ≥ 4 reads in 2, 5 and 9 dpf, respectively, were log transformed and analyzed for expression abundance between genders. The distributions of deviations between genders were analyzed separately in each of 2, 5 and 9 dpf and were approximately normal as exemplified for 9 dpf in Figure S3. Nine sexually differentially expressed miRNAs with deviations between genders ≥ 4 standard deviations were obtained. This stringent criterion corresponds to an FDR of 1.7% by comparison to the number of miRNAs that are expected purely by chance with >4 standard deviations in a normal distribution. The nine differentially expressed miRNAs by gender are presented in Table 3. The miRNAs ranged in expression abundance between genders by approximately two to 10-folds, and were consistent across time points. Three miRNAs were up-regulated in females and six miRNAs in males. Only four of the nine miRNAs had conserved annotated names, i.e. mir-21, pma-mir-4585, bmo-mir-2779 and mir-218. Two miRNAs were independently found in two of the three time points while two others were found in all three. Thus, given the low FDR level and the independent detection of half of the miRNAs at multiple time points with stable differential expression by gender, the nine differentially expressed miRNAs may be considered reliable. Interestingly, most of the up-regulated miRNAs in males were at 2 and 5 dpf and those up-regulated in females were at 5 and 9 dpf.

Analysis of gene targets for differentially expressed miRNAs between genders

The 3' UTR of the 59 genes differentially expressed between genders in the microarray experiment were tested as potential targets for nine miRNAs that were differentially expressed between genders (Table 4). A total of 44 predicted gene targets were identified and presented by gender in Table 4. For each miRNA and gene target combination, concordant and discordant relationship is displayed by plus and minus symbols, respectively. Concordance is called for an up-regulated miRNA and its down-regulated putative gene target. Only one of the miRNAs e.g., pma-mir-4585 that was up-regulated in males, showed significant perfect inverse correlation of expression pattern with its six targeted genes; *cr/20β-hsd*, *psmb8*, *rtn4ip1*, *casp8*, *atp5g3* and an unannotated gene, that were down-regulated in males. Moreover, the first gene is known to activate female determination. This miRNA was up-regulated in male vs. female embryos at both 2 and 5

dpf. At 9 dpf it was up-regulated by only three standard deviations and thus was not marked as differentially expressed at 9 dpf in Table 3.

Analysis of gene hosts for differentially expressed miRNAs between genders

Two of the nine differentially expressed miRNAs were up-regulated in males and highly conserved among vertebrates. Therefore their gene hosts could be determined by across species genomic analysis. mir-21 was identified in *Tubulin Delta 1* (*tubd1*), and mir-218 was found in the 3' region of *developmentally regulated GTP binding protein 1* (*drg1*).

Discussion

The main objective in the present study was to identify genes and miRNAs that were differentially expressed between genders before gonad formation. Since differences in gene expression were previously detected in the bi-potential gonads at 9-10 dpf, we analyzed the embryos at 2, 5 and 9 dpf, which are equivalent to the developmental stages of brain differentiation, hatching and late larva period, respectively [53].

Fifty-nine genes were differentially expressed between genders based on an FDR of 0.05. Correlations of expression patterns between genders were 0.85 and 0.79 between 2 dpf and 5 and 9 dpf, respectively. The correlation between the latter two stages was 0.95. Thus, the somewhat lower correlation of expression between 2 dpf and later embryonic stages may indicate partial transcription at 2 dpf of maternal RNAs stored in oocytes. Most of the detected genes are known to play a role in vertebrate SD and apoptosis. Functional annotation clustering indicated a significant enriched cluster of the immune response containing four genes: *psmb8*, *fcgrt*, *gas7a7* and *zp3*. A recent study detected *sdY* as a sex determining gene in rainbow trout that evolved from an immune related gene [54]. In addition, we found 22 genes that were associated to apoptosis pathways. This is in accordance with previous studies, suggesting that evolutionary conserved genes in the immune system and apoptotic cell death processes may also play a role in this early stage of differentiation and SD [55,56]. Apoptotic pathways are known to be part of sex differentiation in zebrafish [56]. Furthermore, elimination of the Müllerian duct, the primitive female reproductive tract, is triggered by *amh* and mediated also by apoptosis in

mammalian sexual differentiation of male [57]. Thus, apoptosis pathways may be involved in SD or sex differentiation of Nile tilapia.

The *amh* gene, also called Müllerian inhibiting substance, is a member of the TGF-β family that is a key player in cell proliferation, differentiation and apoptosis [58]. It is secreted by Sertoli cells and is responsible for the regression of Müllerian duct during male fetal development in mammals, birds, and reptiles [59]. This gene is also a positional candidate gene due to its location in the central region of the SD QTL on LG23 [16,21]. Fifty-one genes were positional candidates in the 1.47 Mbp critical region of the SD QTL on LG23 in tilapia [21], but only *amh* was differentially expressed in male embryos and testis. A thorough characterization of the gene showed two SNPs in *amh* exon VII among the three unrelated individuals tested, and a novel duplication in males. This unique male-specific copy, denoted *amhy,* has a deletion of 233 bp of the TGF-β domain, and is therefore not capable of encoding the corresponding 86 amino acids. However, a capability of encoding additional upstream 21 amino acids emerged, due to a reading frame shift and disruption of a stop codon. Recently, Y-linked *amh* duplication was identified with a 577 bp insertion in intron 3, and a critical role in SD of Patagonian pejerrey [60]. In addition, SNP in the kinase domain of anti-Müllerian receptor type II *(amhr2)* was found to be associated with SD in fugu [61]. Apparently all other candidate genes that were represented on the array including *cyp19 aromatases*, *dmrt1*, *elavl1*, *gnrh*, *msp*, *sox9*, *sox14*, *lhx9* and *foxl2* were not sexually differentially expressed at the early embryonic development; although they are highly expressed in either ovary or testis of tilapia [41]. This may reflect the pivotal early control of *amh* in SD, and the role of additional downstream genes participating in formation and function of the gonads. However, the mechanism through which a duplicated copy of *amh* lacking its regulatory region may lead to male determination remains unknown. Furthermore, *amhy* could also be an ancient sex determining gene with no effect on male determination in the contemporary SD mechanism of Nile tilapia.

Additional genes that were highly expressed in males were *tspan8, zbed3, wasp1, cyp4f3* and *prkcb*. *tspan8* and *zbed3* were reported as overexpressed in testis, interacting with axin protein thus activating Wnt/β-catenin signaling and TGF-β/BMP pathways, respectively [62,63]. *cyp4f3* was found to be necessary for efficient male mating in *Drosophila melanogaster* and *prkcb* modulates SMAD-dependent TGF-β signaling [64,65].

cr/20ß-hsd was identified as the most significant over expressed gene in females. This gene is known to be part of the oxidoreductase pathway for oocyte maturation preceding the enzymatic activity of *cyp19* (*Cytochrome P450 aromatase*) [66]. *cyp19a1a* was proposed as

the major gene for female determination in zebrafish [56]. Additional overexpressed genes in females such as *rtn4ip1*, *impa1, socs7* and *zp3* are involved in anti-apoptotic activity, embryonic development and fertility, prolactin and Jak-STAT signaling pathway and ovary development, respectively [67-70].

We found several cases of CNV within the set of genes that were differentially expressed between genders. qPCR analysis with gDNA has validated the existence of sexually dimorphic CNV in at least four genes, e.g., *cr/20βhsd*, *tspan8*, *gpa33* and *amh*. All four genes showed increased copy number in the direction of overexpression, in accordance with the reported positive correlation between relative expression level and gene dosage [71]. Five additional differentially expressed genes; *fcgrt*, *socs7*, *urgcp*, *zbed3* and *LOC100690239;* with multiple representations in the genome have been identified. CNV and dosage sensitivity have been hypothesized as evolutionary conserved factors of SD and SD plasticity among related species [72,73]. Hattori et al. [60] suggested that master determinants of SD are predominantly recruited from the duplication of genes involved in the sex differentiation cascade. Our findings, supported by the above studies, indicate that CNV is a common feature of genes participating in SD and may be the alternative genomic structure to sex chromosome systems in fish.

The emerging significance of miRNAs in developmental processes and their ability to regulate large numbers of genes indicate their potential role in determining the onset of SD. We found 704, 668 and 636 miRNAs in tilapia embryos at 2, 5 and 9 dpf. Our findings of nine sexually differentially expressed miRNAs from 2 dpf illustrate their possible role in the early developing embryo. pma-mir-4585 was up-regulated by five and two fold in male *vs*. female embryos at 2 and 5 dpf, respectively, and to a lesser extent of 1.5 fold at 9 dpf. This decay of expression over time may indicate the significance of this miRNA in males soon after fertilization. For pma-mir-4585, an up-regulated miRNA in males, all six predicted target genes, including *cr/20β-hsd* that is known to activate female determination, were down-regulated in males (Table 3). The probability that this should occur by chance is $(1/2)^6$=0.015, thus strengthening the functional targeting of these genes. Only this miRNA showed significantly perfect inverse expression correlation with its targeted genes, in accordance with the expected inhibition of mRNA translation of target genes [74]. Interestingly, this miRNA was firstly identified in sea lamprey (*Petromyzon marinus*) brain, which may be relevant to regulation of SD in fish through its potential activity in the brain.

Two of the nine miRNAs that were up-regulated in males, were conserved among vertebrates and thus allowed a thorough genomic analysis for their host genes; mir-218 was

found to reside on *drg1* gene which had the highest expression in zebrafish testis, as compared to all other tissues tested [50], and mir-21 was localized to the *tubd1* gene which is involved in the elongation of the spermatid through a specialized microtubule system present during reshaping of the sperm head [75]. The function of this gene is related to the Sertoli pathway. Thus, the functions of both host genes in testis is consistent with the findings that miRNAs are usually coordinately coexpressed with their host genes mRNAs, implying that they derive from a common primary transcript [76]. Papagiannakopoulos et al. [77] reported that mir-21 targets several genes in the TGF-β and apoptosis pathways. Further investigation of the functions of the differentially expressed miRNAs in SD of tilapia is warranted.

We found that two thirds of the genes were highly expressed in females, especially at the early developing embryo, i.e. 2 to 5 dpf, and that two thirds of the miRNAs were up-regulated in males at the same period of embryonic development. Histological sex differentiation of the gonads in Patagonian pejerrey showed that the ovary differentiated at 3-4 weeks after hatching, as compared to 5-6 weeks in testis [60]. Based on these observations, it may be postulated that early onset of genes in the female cascade determine female, unless they are down-regulated by miRNAs, thus initiating the male determining pathway.

In summary, this study reports the first discovery of sexually differentially expressed genes, genes enriched for CNV and apoptosis, miRNAs and their predicted gene targets and hosts that are functional from 2 dpf embryos. The experimental workflow used in this study is presented in Figure 7. The gene expression analysis is presented on the left axis along with the miRNAs analysis on the right. Systems biology techniques were used to derive and connect information on genes and miRNAs. The implicated conclusions integrating the two streams of data are presented in the middle of this Figure. *amh* and *cr/20β-hsd* genes may be involved in male and female determination and differentiation, respectively, similar to the proposed SD model for zebrafish [56]. *amhy* segregated in full concordance with the SD-linked marker on LG23 signifying the QTL for SD [21], thus indicating its potential role in SD.

To test the role of potential sex determining genes, miRNA and their predicted gene targets that were found in the current study for tilapia, targeted strategies should be considered, such as (i) mutant detection in candidate genes, as performed in zebrafish [56]; (ii) gene silencing using the TALEN or CRISPR/CAS9 technologies, as applied in tilapia and zebrafish, respectively [78,79]; and (iii) transgenesis which was demonstrated for Nile

tilapia [80,81]. *amh* is a highly prioritized candidate gene for analysis by the variety of suggested methods in order to unravel its potential role in SD of tilapia.

Additional files:

Table S1: Annotations of 59 differentially expressed probes.

Figure S1: PCA analysis for microarray expression data.

Figure S2: False discovery rate for microarray expression data.

Figure S3: Distribution of log transformed miRNAs expression data at 9 dpf.

List of Abbreviations

SD: sex determination, CNV: copy number variation, dpf: days post fertilization.

Competing interests

The authors declare they have no competing interests.

Authors' contribution

Conceived and designed the experiments: OE MR MB GH. Performed the experiments: OE MB AS LD. Collection of tissue samples: TZ OE AS. Analyzed the data: MMG VCC OE MR EF ES JIW. Wrote the manuscript: OE MR GH AS.

Acknowledgements

We thank Prof. Berta Levavi-Sivan (HUJ, Israel) for providing us the ΔXX males developed in her laboratory. We thank Prof. Thomas D. Kocher (UMD, USA) for his support, advice, and critical comments, and for sharing with us genomic resources for Nile tilapia that were developed in his lab. We thank Dr. Federica Di Palma and Dr. Kerstin Lindblad-Toh of the Broad Institute Genome Sequencing Platform and Genome Sequencing and Analysis Program, for making the unpublished genome sequence data for Nile tilapia available. We thank the anonymous reviewers for their constructive comments and suggestions. This is a contribution from the ARO, Institute of Animal Science, Bet Dagan, Israel, No. 681/13. The research was supported by grant No. 801/11 from the Israeli Science Foundation.

References

1. Nelson JS: **Fishes of the World.** *Wiley, New York, NY* 1994:600.

2. Berta P, Hawkins JB, Sinclair AH, Taylor A, Griffiths BL, Goodfellow PN, Fellous M: **Genetic evidence equating SRY and the testis-determining factor**. *Nature* 1990, **348**(6300):448-450.

3. Birchler JA, Bhadra U, Bhadra MP, Auger DL: **Dosage-dependent gene regulation in multicellular eukaryotes: implications for dosage compensation, aneuploid syndromes, and quantitative traits.** *Developmental biology* 2001, **234**(2):275-288.

4. Ferguson MW, Joanen T: **Temperature-dependent sex determination in Alligator mississippiensis.** *Journal of Zoology* 1983, **200**(2):143-177.

5. Mittwoch U: **Sex is a threshold dichotomy mimicking a single gene effect.** *Trends in Genetics* 2006, **22**(2):96-100.

6. Crosetti D, Sola L, Brunner P, Cataudella S: **Cytogenetical characterization of *Oreochromis niloticus*, *O. Mossambicus* and their hybrid.** In *The Second Symposium on Tilapia in Aquaculture ICLARM Department of Fisheries, Bangkok, Thailand and International Center for Living Aquatic Resources Management, Manila, Filipinas*. 1988: 143-151.

7. Baroiller JF, D'Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G: **Tilapia sex determination: Where temperature and genetics meet.** *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* 2009, **153**(1):30-38.

8. Hickling CF: **The Malacca tilapia hybrids.** *Journal of Genetics* 1960, Volume **57**(1):1-10.

9. Pruginin Y, Rothbard S, Wohlfarth G, Halevy A, Moav R, Hulata G: **All-male broods of *Tilapia nilotica* × *T. aurea* hybrids.** *Aquaculture* 1975, **6**(1):11-21.

10. Mair GC, Scott AG, Penman DJ, Beardmore JA, Skibinski DOF: **Sex determination in the genus *Oreochromis*.** *Theoret Appl Genetics* 1991, **82:**144-152.

11. Penman DJ, M.S. Shah, J.A. Beardmore & D.O.F. Skibinski: **Sex ratios of gynogenetic and triploid tilapia.** In *Proceedings of the World Symposium of Selection, Hybridization, and Genetic Engineering in Aquaculture, Vol II, edited by K Tiews Heeneman, Berlin* 1987**:**267–276.

12. Myers JM, Penman DJ, Basavaraju Y, Powell SF, Baoprasertkul P, Rana KJ, Bromage N, McAndrew BJ: **Induction of diploid androgenetic and mitotic gynogenetic Nile tilapia (*Oreochromis niloticus*)**. *Theor Appl Genet* 1995, **90**(2):205-210.

13. Baroiller JF, Chourrout D, Fostier A, Jalabert B: **Temperature and sex chromosomes govern sex ratios of the mouthbrooding cichlid fish *Oreochromis niloticus***. *Journal of Experimental Zoology* 1995, **273**(3):216-223.

14. Lee B-Y, Lee W-J, Streelman JT, Carleton KL, Howe AE, Hulata G, Slettan A, Stern JE, Terai Y, Kocher TD: **A second-generation genetic linkage map of tilapia (*Oreochromis* spp.).** *Genetics* 2005, **170**(1):237-244.

15. Shirak A, Palti Y, Cnaani A, Korol A, Hulata G, Ron M, Avtalion RR: **Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*).** *Journal of Heredity* 2002, **93**(4):270-276.

16. Eshel O, Shirak A, Weller JI, Slossman T, Hulata G, Cnaani A, Ron M: **Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*).** *Animal Genetics* 2011, **42**(2):222-224.

17. Lee B-Y, Penman DJ, Kocher TD: **Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis.** *Animal Genetics* 2003, **34**(5):379-383.

18. Lee B-Y, Hulata G, Kocher T: **Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*).** *Heredity* 2004, **92**(6):543-549.

19. Cnaani A, Tinman S, Avidar Y, Ron M, Hulata G: **Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F_2 tilapia hybrid**. *Aquaculture* 2003, **223**(1):117-128.

20. Cnaani A, Lee B-Y, Zilberman N, Ozouf-Costaz C, Hulata G, Ron M, D'Hont A, Baroiller JF, D'Cotta H, Penman DJ et al: **Genetics of sex determination in tilapiine species**. *Sexual Development* 2008, **2**(1):43-54.

21. Eshel O, Shirak A, Weller JI, Hulata G, Ron M: **Linkage and physical mapping of sex region on LG23 of nile tilapia (*Oreochromis niloticus*).** *G3: Genes|Genomes|Genetics* 2012, **2**(1):35-42.

22. Poonlaphdecha S, Pepey E, Huang SH, Canonne M, Soler L, Mortaji S, Morand S, Pfennig F, Mélard C, Baroiller JF et al: **Elevated AMH gene expression in the brain of male tilapia (Oreochromis niloticus) during testis differentiation**. *Sexual Development* 2011, **5**(1):33-47.

23. Skaar K, Nobrega R, Magaraki A, Olsen L, Schulz R, Male R: **Proteolytically activated, recombinant Anti-Müllerian hormone inhibits androgen secretion, proliferation, and differentiation of spermatogonia in adult zebrafish testis organ cultures**. *Endocrinology* 2011, **152**(9):3527-3540.

24. Weller JI, Ron M: Invited review: **Quantitative trait nucleotide determination in the era of genomic selection.** *Journal of Dairy Science* 2011, **94**(3):1082-1090.

25. Lewis BP, Shih Ih, Jones-Rhoades MW, Bartel DP, Burge CB: **Prediction of mammalian microRNA targets.** Cell 2003, **115**(7):787-798.

26. Huang CW, Li YH, Hu SY, Chi JR, Lin GH, Lin CC, Gong HY, Chen JY, Chen RH, Chang SJ et al: **Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (*Oreochromis niloticus*).** Journal of Animal Science 2012, **90**(12):4266-4279.

27. Yan B, Guo J-T, Zhao L-H, Zhao J-L: **MicroRNA expression signature in skeletal muscle of Nile tilapia.** *Aquaculture* 2012, **364–365**(0):240-246.

28. Morgan C, Bale T: **Sex differences in microRNA regulation of gene expression: no smoke, just miRs.** *Biology of Sex Differences* 2012, **3**(1):22.

29. Mishima T, Takizawa T, Luo S-S, Ishibashi O, Kawahigashi Y, Mizuguchi Y, Ishikawa T, Mori M, Kanda T, Goto T et al: **MicroRNA (miRNA) cloning analysis reveals sex differences in mirna expression profiles between adult mouse testis and ovary**. *Reproduction* 2008, **136**(6):811-822.

30. Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, Lao K, O'Carroll D, Das PP, Tarakhovsky A, Miska EA et al: **MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis.** *Plos One* 2008, **3**(3):e1738.

31. Bannister SC, Tizard MLV, Doran TJ, Sinclair AH, Smith CA: **Sexually dimorphic microRNA expression during chicken embryonic gonadal development.** *Biology of Reproduction* 2009, **81**(1):165-176.

32. Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, Baskerville S, Hammond SM, Bartel DP, Schier AF: **MicroRNAs regulate brain morphogenesis in zebrafish**. *Science* 2005, **308**(5723):833-838.

33. D'Cotta H, Fostier A, Guiguen Y, Govoroun M, Baroiller J-F: **Aromatase plays a key role during normal and temperature-induced sex differentiation of tilapia *Oreochromis niloticus***. *Molecular Reproduction and Development* 2001, **59**(3):265-276.

34. Devlin RH, Nagahama Y: **Sex determination and sex differentiation in fish: An overview of genetic, physiological, and environmental influences**. *Aquaculture* 2002, **208**(3–4):191-364.

35. Ijiri S, Kaneko H, Kobayashi T, Wang D-S, Sakai F, Paul-Prasanth B, Nakamura M, Nagahama Y: **Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*.** *Biology of Reproduction* 2008, **78**(2):333-341.

36. Rougeot C, Prignon C, Ngouana Kengne CV, Mélard C: **Effect of high temperature during embryogenesis on the sex differentiation process in the Nile tilapia, *Oreochromis niloticus*.** *Aquaculture* 2008, **276**(1–4):205-208.

37. Mair GC, Abucay JS, Beardmore JA, Skibinski DOF: **Growth performance trials of genetically male tilapia (GMT) derived from YY-males in *Oreochromis* *niloticus* l.**: On station comparisons with mixed sex and sex reversed male populations. *Aquaculture* 1995, **137**(1–4):313-323.

38. Lee B-Y, Howe A, Conte M, D'Cotta H, Pepey E, Baroiller J-F, di Palma F, Carleton K, Kocher T: **An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags.** *BMC Genomics* 2010, **11**(1):278.

39. MIRA: **http://www.chevreux.org/projects_mira.html.**

40. Browzer EG: **http://www.ensembl.org/index.html**. February 2014.

41. Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovsky R, Zilberman N, Kocher TD, Hulata G, Ron M: ***AMH* and *DMRT2* genes map to tilapia (*Oreochromis* spp.) linkage**

group 23 within quantitative trait locus regions for sex determination. *Genetics* 2006, **174**(3):1573-1581.

42. Bioconductor: http://www.bioconductor.org/packages/2.8/bioc/html/limma.html. Version: 2.8.

43. Omnibus GE: **http://www.ncbi.nlm.nih.gov/geo/.**

44. Software P: **http://www.partek.com/?q=software.**

45. Gene Functional Classification Tool DAVID Bioinformatics Resources 6.7 NN: **http://david.abcc.ncifcrf.gov/gene2gene.jsp**.

46. Primer3Plus: **http://www.bioinformatics.nl/cgi-bin/primer3plus/primer3plus.cgi/.**

47. Abyzov A, Mariani J, Palejev D, Zhang Y, Haney MS, Tomasini L, Ferrandino AF, Belmaker LAR, Szekely A, Wilson M: **Somatic copy number mosaicism in human skin revealed by induced pluripotent stem cells.** *Nature* 2012, **492**(7429):438-442.

48. Staden R, Beal K, Bonfield J: The staden package, 1998. In: **Bioinformatics Methods and Protocols.** Edited by Misener S, Krawetz S, vol. **132**: *Humana Press*; 1999: 115-130.

49. Friedländer MR, Chen W, Adamidi C, Maaskola J, Einspanier R, Knespel S, Rajewsky N: **Discovering microRNAs from deep sequencing data using mirdeep.** Nature *Biotechnology* 2008, **26**(4):407-415.

50. BioGPS: **http://biogps.org/#goto=welcome.**

51. Krüger J, Rehmsmeier M: RNAhybrid: **MicroRNA target prediction easy, fast and flexible.** *Nucleic Acids Research* 2006, **34**(suppl 2):W451-W454.

52. Enright AJ, John B, Gaul U, Tuschl T, Sander C, Marks DS: **MicroRNA targets in Drosophila.** *Genome biology* 2004, **5**(1):R1-R1.

53. Fujimura K, Okada N: **Development of the embryo, larva and early juvenile of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae). Developmental staging system.** *Development, growth & differentiation* 2007, **49**(4):301-324.

54. Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, Klopp C, Cabau C, Bouchez O, Fostier A, Guiguen Y: **An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss.*** *Current Biology* 2012, **22**(15):1423-1428.

55. Opferman J: **Apoptosis in the development of the immune system**. *Cell Death & Differentiation* 2007, **15**(2):234-242.

56. Rodríguez-Marí A, Cañestro C, BreMiller RA, Nguyen-Johnson A, Asakawa K, Kawakami K, Postlethwait JH: **Sex reversal in zebrafish *fancl* mutants is caused by tp53-mediated germ cell apoptosis.** *Plos Genet* 2010, **6**(7):e1001034.

57. Nef S, Parada LF: **Hormones in male sexual development.** *Genes & development* 2000, **14**(24):3075-3086.

58. Kubiczkova L, Sedlarikova L, Hajek R, Sevcikova S: **Tgf-beta - an excellent servant but a bad master.** *Journal of Translational Medicine* 2012, **10**(1):183.

59. Josso N, Cate R, Picard J, Vigier B, Di Clemente N, Wilson C, Imbeaud S, Pepinsky R, Guerrier D, Boussin L: **Anti-mullerian hormone: the Jost factor.** *Recent progress in hormone research* 1992, **48**:1-59.

60. Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, Sakamoto T, Fernandino JI, Somoza GM, Yokota M, Strüssmann CA: A **Y-linked anti-müllerian hormone duplication takes over a critical role in sex determination.** *Proceedings of the National Academy of Sciences* 2012, **109**(8):2955-2959.

61. Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, Mizuno N, Fujita M, Suetake H, Suzuki S, Hosoya S et al: **A trans-species missense SNP in Amhr2 is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (*Fugu*).** *Plos Genet* 2012, **8**(7):e1002798.

62. Lardenois A, Chalmel F, Barrionuevo F, Demougin P, Scherer G, Primig M: **Profiling spermatogenic failure in adult testes bearing *SOX9*-deficient sertoli cells identifies genes involved in feminization, inflammation and stress.** Reproductive Biology and *Endocrinology* 2010, **8**(1):154.

63. Chen T, Li M, Ding Y, Zhang LS, Xi Y, Pan WJ, Tao DL, Wang JY, Li L: **Identification of zinc-finger BED domain-containing 3 (*Zbed3*) as a novel Axin-interacting protein that activates Wnt/beta-catenin signaling.** *The Journal of biological chemistry* 2009, **284**(11):6683-6689.

64. Fujii S, Toyama A, Amrein H: **A male-specific fatty acid ω-hydroxylase, *sxe1*, is necessary for efficient male mating in drosophila melanogaster.** *Genetics* 2008, **180**(1):179-190.

65. Yakymovych I, Dijke P, Heldin CH, Serhiy S: **Regulation of Smad signaling by protein kinase C**. *The FASEB Journal* 2001, **15:**553-555.

66. Senthilkumaran B, Yoshikuni M, Nagahama Y: **A shift in steroidogenesis occurring in ovarian follicles prior to oocyte maturation.** *Molecular and Cellular Endocrinology* 2004, **215**(1–2):11-18.

67. Watari A, Yutsudo M: **Multi-functional gene asy/nogo/rtn-x/rtn4: Apoptosis, tumor suppression, and inhibition of neuronal regeneration.** *Apoptosis* 2003, **8**(1):5-9.

68. Cryns K, Shamir A, Van Acker N, Levi I, Daneels G, Goris I, Bouwknecht JA, Andries L, Kass S, Agam G, et al: ***IMPA1* is essential for embryonic development and lithium-like pilocarpine sensitivity.** *Neuropsychopharmacology : official publication of the American College of Neuropsychopharmacology* 2008, **33:**674-684.

69. Database KP: **http://www.genome.jp/kegg/pathway.html.**

70. Garcia-Ortiz J, Pelosi E, Omari S, Nedorezov T, Piao Y, Karmazin J, Uda M, Cao A, Cole S, Forabosco A et al: ***FOXL2* functions in sex determination and histogenesis throughout mouse ovary development.** *BMC Developmental Biology* 2009, **9**(1):36.

71. Henrichsen CN, Chaignat E, Reymond A: **Copy number variants, diseases and gene expression.** *Human Molecular Genetics* 2009, **18**(R1):R1-R8.

72. Schartl M: **Sex chromosome evolution in non-mammalian vertebrates.** *Curr Opin Genet Dev* 2004, **14**(6):634-641.

73. Volff JN, Nanda I, Schmid M, Schartl M: **Governing sex determination in fish: Regulatory putsches and ephemeral dictators.** *Sexual development : genetics, molecular biology, evolution, endocrinology, embryology, and pathology of sex determination and differentiation* 2007, **1:**85-99.

74. Guo H, Ingolia NT, Weissman JS, Bartel DP: **Mammalian microRNAs predominantly act to decrease target mrna levels.** *Nature* 2010, **466**(7308):835-840.

75. Smrzka OW, Delgehyr N, Bornens M: **Tissue-specific expression and subcellular localisation of mammalian δ-tubulin.** *Current Biology* 2000, **10**(7):413-416.

76. Baskerville S, Bartel DP: **Microarray profiling of microRNAs reveals frequent coexpression with neighboring miRNAs and host genes.** RNA 2005, 11(3):241-247.

77. Papagiannakopoulos T, Shapiro A, Kosik KS: **MicroRNA-21 targets a network of key tumor-suppressive pathways in Glioblastoma cells.** *Cancer Research* 2008, **68**(19):8164-8172.

78. Li MH, Yang HH, Li MR, Sun YL, Jiang XL, Xie QP, Wang TR, Shi HJ, Sun LN, Zhou LY: **Antagonistic roles of *Dmrt1* and *Foxl2* in sex differentiation via estrogen production in tilapia as demonstrated by TALENs.** *Endocrinology* 2013, **154**(12):4814-4825.

79. Hwang WY, Fu Y, Reyon D, Maeder ML, Tsai SQ, Sander JD, Peterson RT, Yeh JJ, Joung JK: **Efficient genome editing in zebrafish using a CRISPR-Cas system.** *Nature biotechnology* 2013, **31**(3):227-229.

80. Fujimura K, Kocher TD: **Tol2-mediated transgenesis in tilapia (*Oreochromis niloticus*).** *Aquaculture* 2011, **319**(3–4):342-346.

81. Golan M, Levavi-Sivan B: **Social dominance in tilapia is associated with gonadotroph hyperplasia.** *General and Comparative Endocrinology* 2013, **192**(0):126-135.

Figure legends

Figure 1 Tilapia gene expression eArray design. The number of probes originating from each one of three types of bioinformatic analyses is presented in different colors. Redundant sequences from different analyses were removed.

Footnote

Genome: Nile tilapia genome first draft sequences, SD: sex determination, FISH ORF: open reading frames of publicly available fish proteins (zebrafish, tetraodon, fugu, and stickleback).

Figure 2 Clustering of differentially expressed genes by gender and dpf. Fifty nine differentially expressed genes between males and females at 2, 5 and 9 dpf, at false discovery rate of $p<0.05$, were clustered. The Z-score represents the number of standard deviation units of genes' expression value from the mean (defined as 0). Blue and yellow colors indicate up and down expression, respectively. Vertical bars represent the clustering of genes based on expression profiles. *urgcp* and *zbed3* genes appear more than once and represent different probes at multiple locations in the genome. Probe numbers begin with the prefix "CUST_". Annotations of probes are presented in Table S1. LOC numbers are NCBI IDs.

Figure 3 *amh* and *cr/20β-hsd* normalized relative expression. Gene expression by qPCR is presented for males (blue) and females (purple) in 2 to 9 dpf embryos (A), and in brain, testis/ovary and liver at 75 dpf (B). Deviation bars represent standard errors and asterisks represent the level of significance for sex-specific expression differences: ** $p \leq 0.001$.

Figure 4 *tspan8, gpa33* and *cr/20β-hsd* copy number variation. Normalized relative genomic quantity by qPCR is presented for males (blue) and females (purple) for analysis of copy number variation. A and B represent crosses involving different dams. Deviation bars represent standard errors. Asterisks represent the level of significance for sex-specific expression differences: * $p \leq 0.05$; ** $p \leq 0.01$ and *** $p<0.001$.

Figure 5 Identification of Y-linked *amh* duplication. (A) Schematic illustration of the full length *amh* gene. Lines shaded with green, introns; Red boxes, exons; the Roman numerals outside the boxes indicate exon number. Four sets of PCR genomic fragments are presented under the respective parts of the gene for XX, XY and YY DNA samples. (B) DNA sequence traces of *amh* and *amhy*

exon VII of three unrelated individuals: XX female, XY and YY males (GenBank: HG518783-6). Capital letters under the traces denote the deduced capable of encoding amino acids. Purple arrow, SNP in nucleotide position 1,274 of the full length *amh* gene in YY individual; green arrow, A>G substitution in nucleotide 1,275 of *amhy* between XY and YY individuals; red arrows, deletion starts in *amhy* from nucleotide position 1,403. (C) PCR for exon VII from cDNA of male and female 3 and 6 dpf embryos, brain, liver, and gonads.

Figure 6 Genomic sequences and predicted polypeptides of *amh* and *amhy* genes. The coding region of the 7th exon of the *AMH* gene is aligned with sequences of XX-female *amh* and YY-male *amhy*. Sequences derived from GenBank entries were aligned (Accession Nos. XM_0034513050, HG518787, HG518785 for *amh* reference gene, female *amh* and male *amhy*, respectively). Above the sequences, Roman numerals label the genomic element of the gene. Asterisks below the sequences denote identical residues in all three sequence submissions. Dashes mark gaps introduced by the alignment program. The shadowed regions localize the TGF-β domain. Numbers indicating the position of the last residue within the GenBank entry are added at right ends of rows. (A) Genomic sequences: The end of intron 6 is shown in lower-case italic letters and the last two bases of the acceptor splice site (ag) are in bold type. The in frame stop codons (TAG or TAA) are in bold and underlined type. (B) Predicted polypeptides: Below the sequences, conserved substitutions are indicated by two dots; and semi-conserved substitutions are indicated by dots. Cysteine residues that form disulfide bonds according to TGF-β family signature (PROSITE PDOC00223) are in white against purple background and this signature layout is delineated above the TGF-β domain sequence.

Figure 7 Experimental workflow for detection of genes and miRNAs affecting SD and sex differentiation. The differentially expressed genes and miRNAs between genders and their integrative role in SD and sex differentiation are displayed. The gene expression analysis is presented on the left axis, and the miRNAs analysis on the right. The implicated conclusions integrating the two streams of data are presented in the middle.

Footnote

SD: Sex determination, mir: miRNA precursor, miR: expressed 3p/5p mature sequence from miRNA precursors, *amh*: Anti-Müllerian hormone, *cr/20β-hsd*: carbonyl reductase 20 beta-hydroxysteroid dehydrogenase, TGF-β: transforming growth factor beta, FDR: false discovery rate, CNV: copy number variation.

→ Experimental flow; —◆ Implicated from the results; ___ Implicated from the literature;
⊣ Negative regulation

Table 1 Primers for gene expression validation with qPCR and *amh* PCR and sequence analysis.

Gene symbol /probe	Description	Accession number	Forward and reverse primers (from 5' to 3')
tspan8	Tetraspanin-8	XM_003448079	TGTATGTTGTGGAATAGGCATCA GGTGATTTGTAAAGCTGTTTCG
fcgrt	Major histocompatibility complex class I-related gene protein-like	XP_003459279	TGGTGTGAGCAAGGACTTCAT ACATGAACCAAAGGACTGTAAACT
cr/20ß-hsd	Carbonyl reductase-like 20 beta-hydroxysteroid dehydrogenase	XM_003438355	CCAAAATTGTTCGTTTTATTCTCG TTTCATTTTGATGCGTTCCA
gpa33	Cell surface A33 antigen	XM_005466820	AATGTCCAAAAGCCAACCTAAA TACTACATCTGCACCTCGGAGA
rtn4Ip11	Reticulon-4-interacting protein 1 homolog, mitochondrial	XM_003459954	GCATGTCAAGGCATCAAATAAA CCTCTGCTGGTTGTAAATGTGA
26098	Hypothetical protein	XM_003455197	TTCCTGAAGACAGTACAGTACAAAA CGAATTCTTCTGGTCAAGTTCTTC
zp3	Hypothetical protein	XM_003439075	TGTCTGTAACTACTCATTTGGATCA CCATTTTACAGATCCAACTTTCC
gapdh	Glyceraldehyde-3-phosphate dehydrogenase	XM_003452690	GGCATCGTGGAAGGTCTCAT CATTTTACCAGAGGGCCCGT
rpp30	Ribonuclease P protein subunit p30	XM_003449476	CCCGACTCCTATCAACGAAC AAAGTGACTCGCGTCTGACA
amh	Anti-Müllerian hormone[1]	DQ257619.1	TTCTTATCGCTCCGACTTCTTC TAGGGCTGGTTGATATGGAATC
	Anti-Müllerian hormone - exon VII[2]	XM_003451305	AGCAGCTCTAGCGGCATCCACA TGTGTTTTCTTTCTGCGTCCGCCA
	1[3]: 5' UTR , exon I, intron 1, exon II	ENSONIG00000004781	AGAGGAGTCATCAGTCCAAAGC AGATGTCCTCCACGAAGCAT
	2: exon II, intron II, exon III	ENSONIG00000004781	AAGACCCCATCATCACCATC TTGTCTGAGCCGTAATCTGC
	3: exon IV, intron III, exon V, intron IV, exon VI	ENSONIG00000004781	GGAAAATCATCAGAGGGGAGT CTGCCGACTTCAGAACTTTT
	4: intron VI, exon VII , 3' UTR	ENSONIG00000004781	CGGTCCCAGTGACCTATGAG AAGTACACGTGGTGTATTGTAATTGA
	5: 3' UTR	ENSONIG00000004781	CCCCAGCATTTATAACTTTCACA CCTGCCTCAAGTATGCCTTT
	LM[4]: intron VI, exon VII	ENSONIG00000004781	TGTGTTTTCTTTCTGCGTCCGCCA AGCAGCTCTAGCGGCATCCACA

[1]qPCR validation of microarray
[2]PCR of cDNA
[3]set of 5 primer pairs for sequencing
[4]Linkage mapping

Table 2 Gene expression validation of microarray results. Mean ratio of female to male gene expression by qPCR from cDNA of embryos from 2 to 9 dpf is presented. The qPCR analysis of genomic DNA (gDNA) was applied for indication of copy number variation.

				qPCR				
Over expressed in	**Gene symbol**	**Microarray**		**cDNA**		**Multiple Tm peaks[3]**	**gDNA**	
		FC[1]	**p-value**	**FC**	**p-value**		**FC**	**p-value**
Male	*amh*	0.5	6×10^{-9}	0.27	0.002		0.75	7.4×10^{-4}
	tspan8	0.2	4.4×10^{-16}	0.0001	5×10^{-11}		0.0013	1.2×10^{-4}
	fcgrt	0.5	6.5×10^{-8}	ND[2]		√	ND	
Female	*cr/20ß-hsd*	3.5	1.0×10^{-17}	352	2.8×10^{-11}	√	669	0.002
	gpa33	3.4	6.0×10^{-9}	88.7	2.6×10^{-5}	√	15.8	3.3×10^{-6}
	rtn4Ip1	2.5	1.0×10^{-17}	2.1	0.0001		ND	
	CUST_26098	4.2	1.6×10^{-7}	4.46	0.004		ND	
	zp3	2.6	3.0×10^{-9}	2	0.004		ND	

[1]FC: fold change of female to male expression: LS means for Microarray and means for qPCR based on cDNA of embryos from 2 to 9 dpf.

[2]ND: not determined

[3]Multiple TM peaks resulting from melting curve analysis

Table 3 miRNA differential expression between genders. miRNA are displayed with female to male normalized expression at 2, 5 and 9 dpf with ≥ four standard deviations of fold change expression in a normal distribution.

#miRNA[1]	dpf / Over expressed in	2	5	9
miR-218	Male	0.4		
pma-miR-4585		0.2	0.5	
oni-miR-E224-		0.1	0.3	0.4
oni-miR-E622-		0.4	0.5	0.6
oni-miR-E218-			0.5	
mir-21			0.6	
oni-miR-E255-	Female	3.8		
oni-miR-E304-			2	2
bmo-miR-2779				1.7

[1]miR: expressed 3p/5p mature sequence from miRNA precursors

Table 4 Differential expression between genders of miRNA and their gene targets. Concordant and discordant differential expression is displayed for nine miRNAs with their forty four differentially expressed between genders predicted gene targets by a false discovery rate of p<0.05.

Gene \ miRNA[1]			Overexpressed in[2]								
			Male						Female		
Probe[3]	Gene symbol[4]	Down-regulated in	miR-218	oni-miR-E218	oni-miR-E224	miR-21	pma-miR-4585	oni-miR-E622	bmo-miR-2779	oni-miR-E255	oni-miR-E304
42507	*fcgrt*	Female			−	−		−			
86	*amh*		−		−			−			+
41041	*urgcp*			−	−	−		−			
38100	*ctnnd2b*		−		−	−					
25051	*g3s1a18*										
25051	*g3s7a7*				−						
21001	*prkcb*										
41366	*zbed3*		−						+		
41366	*7813*		−		−						
41366	*5031*		−	−							
41052	*urgcp*		−		−	−		−		+	
21060	*cyp4f3*		−	−	−			−			
14085	*wasp1*					−					
22935	*urgcp*				−	−		−			+
27670	*zebd3*		−						+		
25225	*cdh1*										
39928	*gpr144*				−	−		−		+	+
16862	*ror1*		−					−			
28545	*fam131c*	Male	+							−	
10109	*eci1*			+		+					
5193	*itgb3a*					+					
13137	*timm50*		+	+	+	+					
555	*fndc7*					+		+			
35393	*nkain1*							+			
6825	*kiaa1468*				+						
7935	*rasgef1b*		+			+		+			
452	*higd1a*				+	+		+			
5723	*sult3*							+			
30073	*casp8*						+				−
23873	*orni-dba*							+			
20878	*hspa5*				+						
31928	*socs7*							+			
9238	*atp5g3*				+		+				
3491	*LOC100703324*		+		+						
35899	*LOC100707490*		+				+				
800	*mrpl34*					+		+			
23235	*psmb8*						+				
13679	*ankrd22*					+		+		−	
24221	*LOC100690239*							+			
5738	*sult3st1*							+			
25471	*rtn4ip1*						+				
1524	*zp3*							+			
9317	*gpa33*				+			+			
1135	*cr/20β-hsd*		+				+				
Rate of concordant target sites			6/15	2/5	7/17	8/15	6/6*	13/21	2/2	2/4	3/4

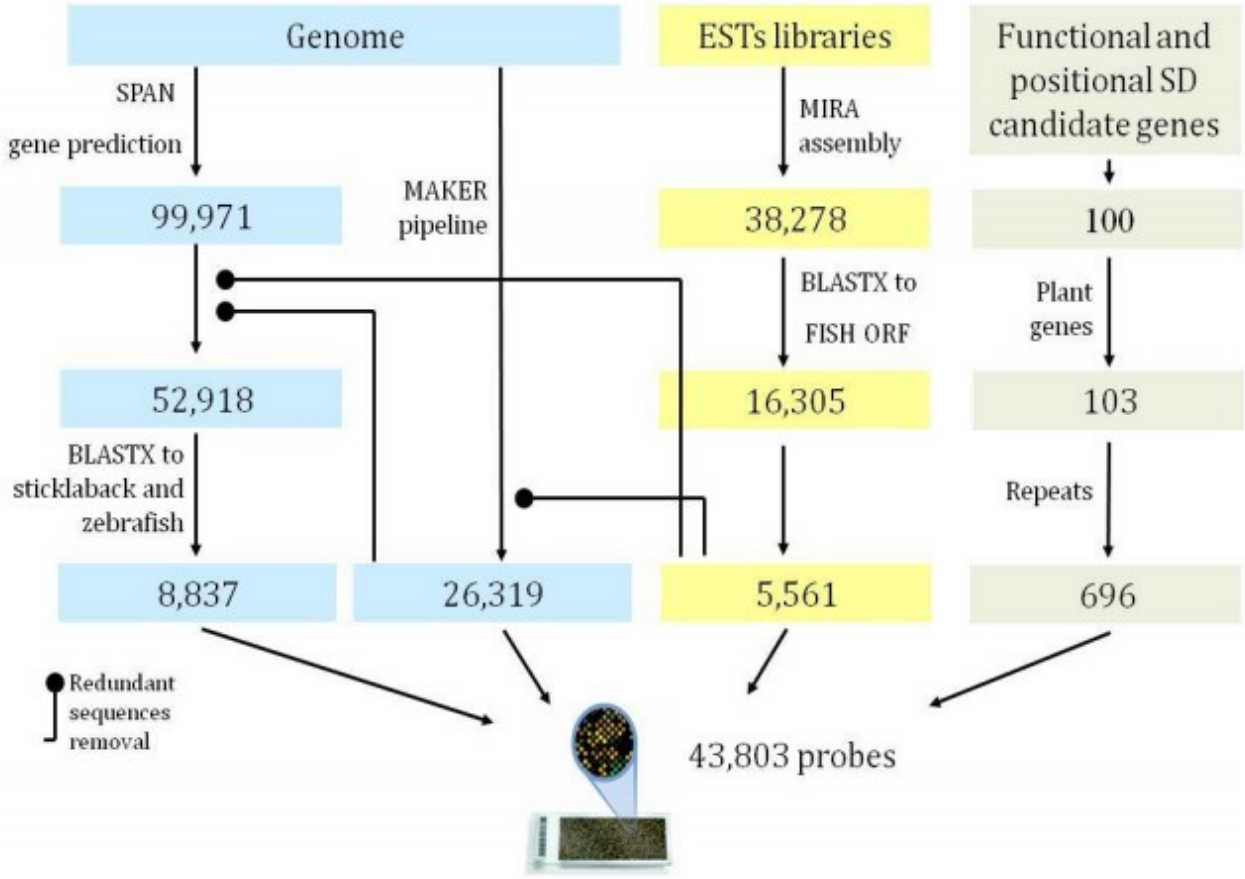
[1]miR: expressed 3p/5p mature sequence from miRNA precursors

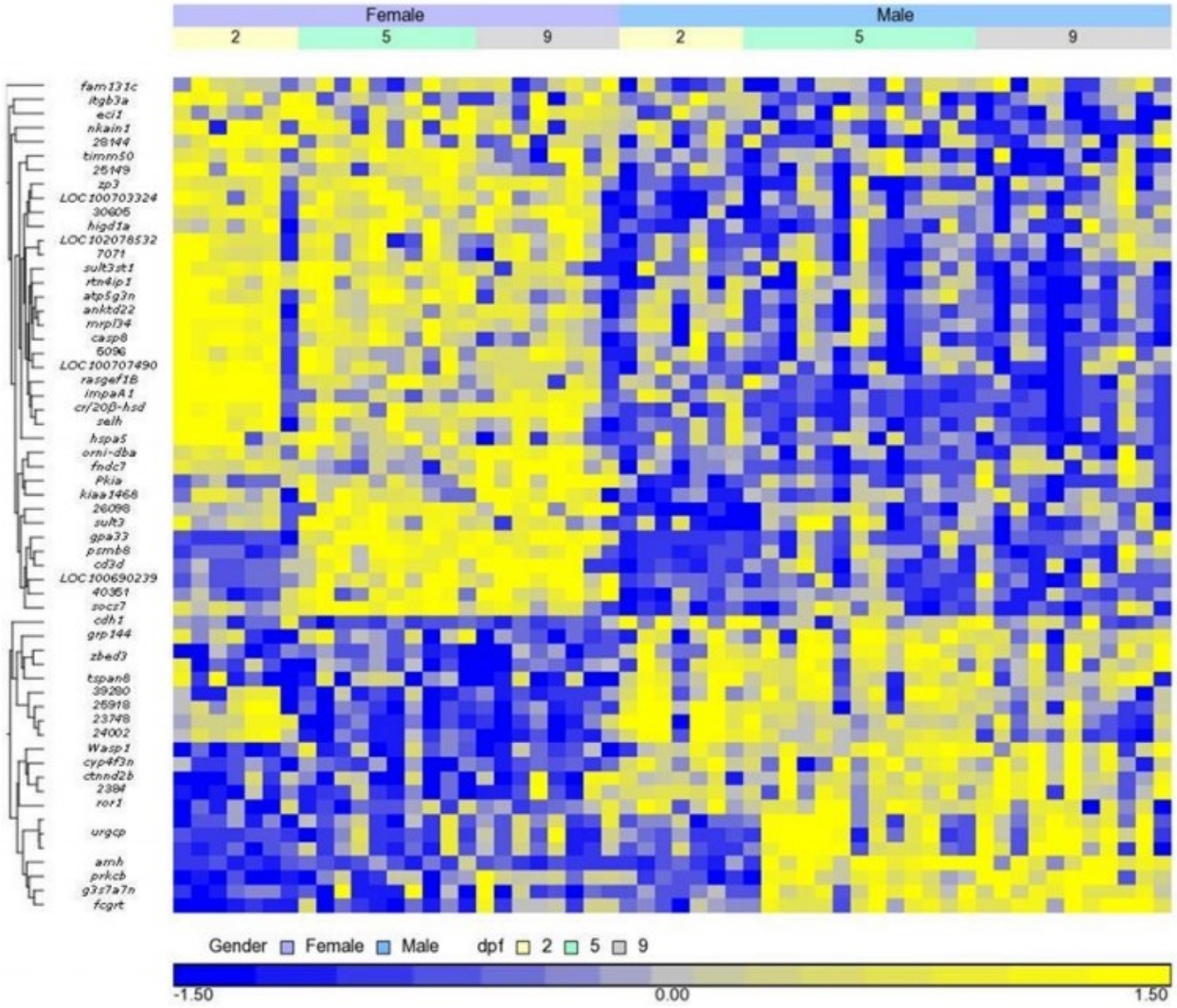
[2]Target recognition sites for miRNAs are represented by concordant (+) or discordant (-) differential expression, respectively, assuming that up-regulated miRNAs down-regulate their target sites [104].

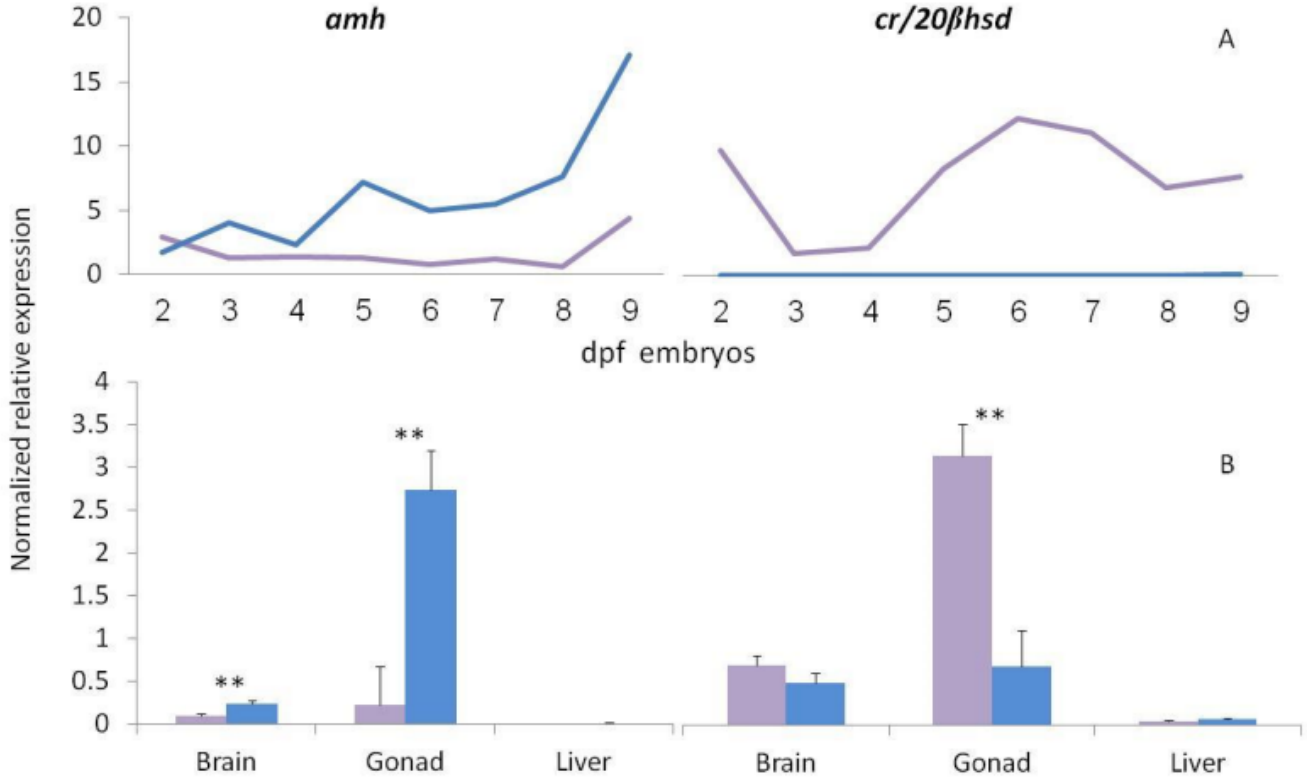
[3]Probe numbers begin with the prefix "CUST_".

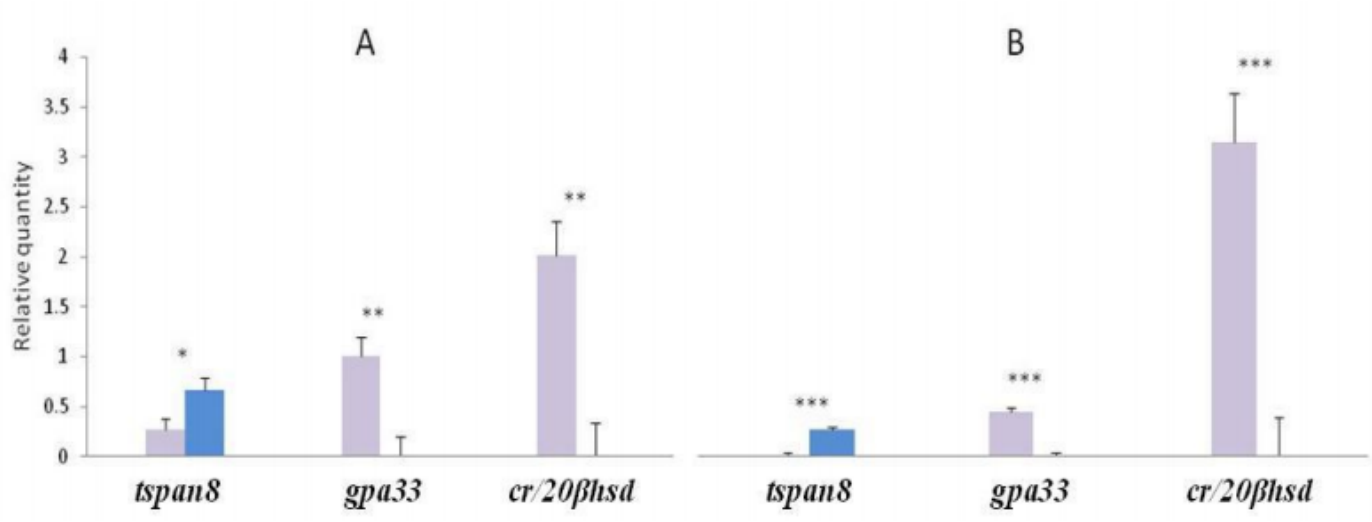
[4]Gene symbol based on NCBI, ENSEMBL or tilapia LOC no. *urgcp* and *zbed3* genes represent different probes at multiple locations in the genome.

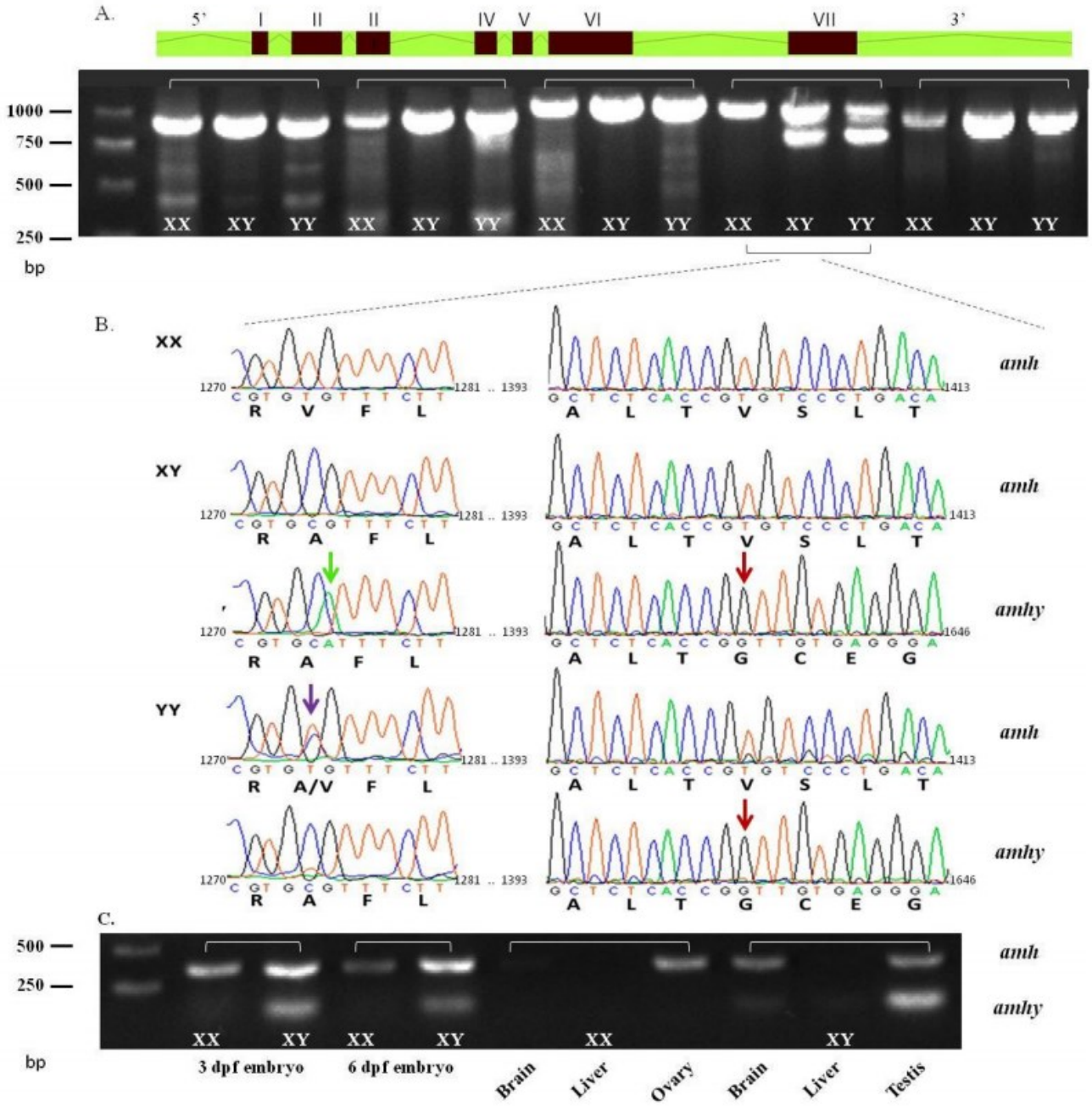
[*]The probability that the rate of concordant target sites should occur by chance is: $(1/2)^6=0.015$

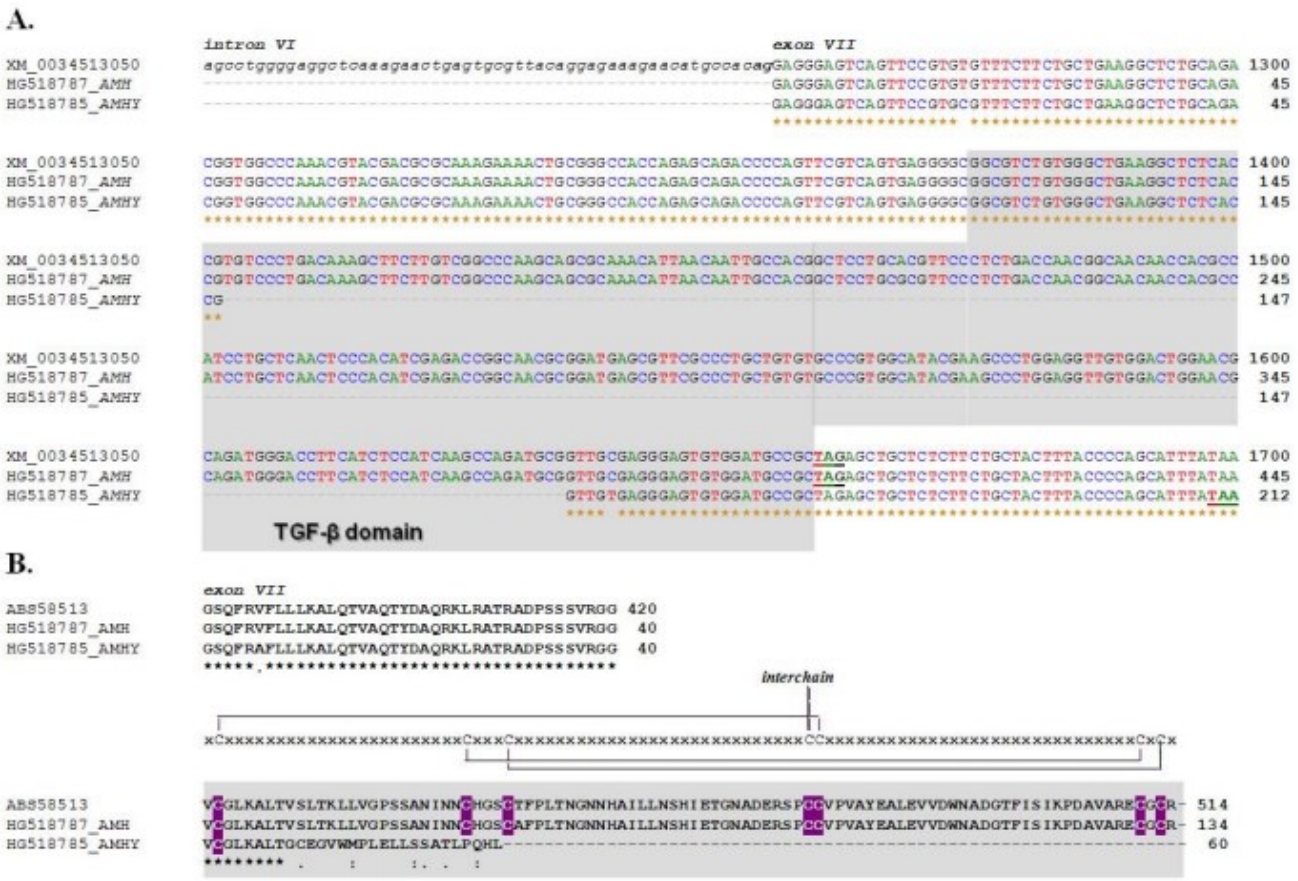

Genome
ESTs libraries
Functional and positional SD candidate genes
SPAN
gene prediction
MIRA
assembly
99,971
MAKER
pipeline
38,278
100
BLASTX to
FISH ORF
Plant
genes
52,918
16,305
103
BLASTX to
sticklaback and
zebrafish
Repeats
8,837
26,319
5,561
696
Redundant
sequences
removal
43,803 probes


Female
Male
2
5
9
2
5
9
fam131c
rtgb3a
eci1
nkain1
28144
timm50
25149
zp3
LOC100703324
30605
higd1a
LOC102076532
7071
sult3st1
rtn4ip1
atp5g3n
anktd22
mrpl34
casp8
5096
LOC100707490
rasgef1B
impaA1
cyl20β-hsd
selh
hspa5
orni-dba
fndc7
Pkia
kiaa1468
26098
sult3
gpa33
psmb8
cd3d
LOC100690239
40351
socs7
cdh1
grp144
zbed3
tspan8
39280
25918
23748
24002
Wasp1
cyp4f3n
ctnnd2b
2384
ror1
urgcp
amh
prkcb
g3st3a7n
fcgrt
Gender Female Male dpf 2 5 9
-1.50
0.00
1.50


amh
cr/20βhsd
A
20
15
10
5
0
2 3 4 5 6 7 8 9
dpf embryos
Normalized relative expression
4
3.5
3
2.5
2
1.5
1
0.5
0
**
**
**
B
Brain
Gonad
Liver
Brain
Gonad
Liver


A
B
Relative quantity
4
3.5
3
2.5
2
1.5
1
0.5
0
*
**
**
***
***
***
tspan8
gpa33
cr/20βhsd
tspan8
gpa33
cr/20βhsd


A.
5'
I
II
II
IV
V
VI
VII
3'
1000
750
500
250
bp
XX XY YY XX XY YY XX XY YY XX XY YY XX XY YY
B.
XX
XY
YY
amh
amhy
R V F L
R A F L
R A/V F L
A L T V S L T
A L T G C E G
C.
500
250
bp
XX XY XX XY XX XY
3 dpf embryo
6 dpf embryo
Brain
Liver
Ovary
Brain
Liver
Testis


A.
intron VI
exon VII
XM_0034513050
HG518787_AMH
HG518785_AMHY
TGF-β domain
B.
exon VII
ABS58513
HG518787_AMH
HG518785_AMHY
interchain

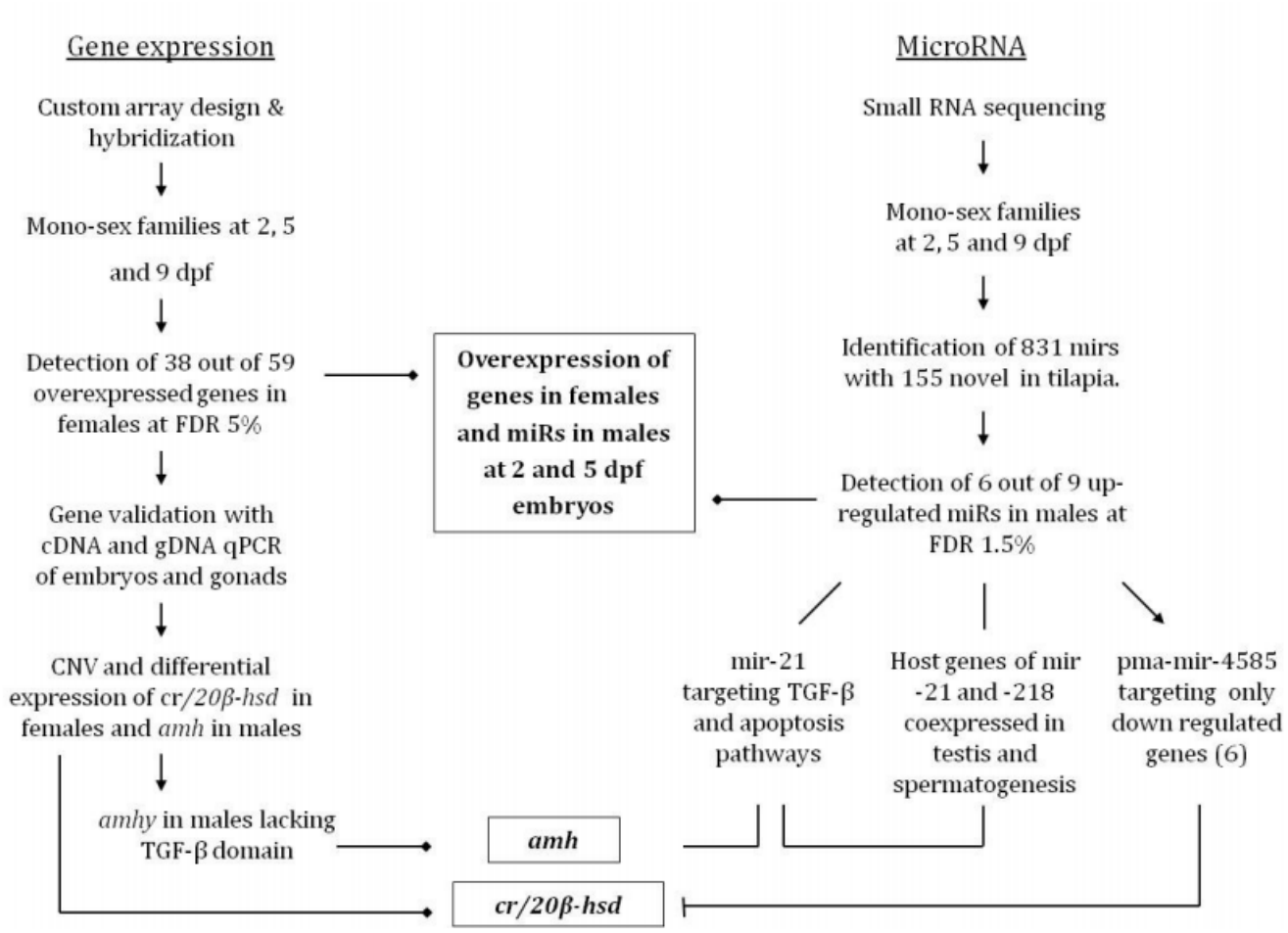

Gene expression
Custom array design & hybridization
Mono-sex families at 2, 5 and 9 dpf
Detection of 38 out of 59 overexpressed genes in females at FDR 5%
Gene validation with cDNA and gDNA qPCR of embryos and gonads
CNV and differential expression of cr/20β-hsd in females and amh in males
amhy in males lacking TGF-β domain
MicroRNA
Small RNA sequencing
Mono-sex families at 2, 5 and 9 dpf
Identification of 831 mirs with 155 novel in tilapia.
Detection of 6 out of 9 up-regulated miRs in males at FDR 1.5%
Overexpression of genes in females and miRs in males at 2 and 5 dpf embryos
mir-21 targeting TGF-β and apoptosis pathways
Host genes of mir -21 and -218 coexpressed in testis and spermatogenesis
pma-mir-4585 targeting only down regulated genes (6)
amh
cr/20β-hsd

Table S1. Characterization and description of 59 differentially expressed genes in the microarray experiment at false discovery rate of p<0.05.

Array probe	Probe sequence
CUST_12128_PI426896222	AATGATAAAATAGTTTGTGTCCAAGAGAGGTGAAGCGAAACAGCTTTACAAATCACCCAG
CUST_2384_PI426896222	CTGTACATTTTTGGATGGATCTGCTTAAGTTCTGCGATGTATTTTGGTACATTGTACAAT
CUST_23748_PI426896222	TTCTGCGTCTCGTCTGAAGTACGGGAGTCAATTAAAGTAACAAAAAGAATTTGGGTCTAA
CUST_25918_PI426896222	TGTTCAGTTCTAAAAGTGGACCATATCCACTCTGTGTGCTGATCAGGGCATTGTCCTGAA
CUST_39280_PI426896222	CCCGTAGTCTTTTACAATACAATGATCAAGCCATTACGTTTATCTTTTGGTACACAATTG
CUST_42507_PI426896222	TGGAGGAGGTACGACTGTGTGTTTCAGATCTCTGGTGTGAGCAAGGACTTCATCATTAAA
CUST_86_PI426901304	GTGTGATTCCATATCAACCAGCCCTAAATACTGTGGAAATCCTATAGTGTTGCATAATAA
CUST_41041_PI426896222	TAGTCACAGTTCTTGGAGTTCAGAGCACAGGAAAGTCCACTCTCCTTAACACCATGTTTG
CUST_24002_PI426896222	TGCGTTAAAGCAGGAAACGTGTTACTATTAGTTAGAAAGCGGGGTTTAAGGCCTCGTCTC
CUST_38100_PI426896222	TCCCTATGGACCCCACTTTTCTCTCTTGTTTAAACACTCAAAGTTCAAACCTTCATAGAA
CUST_25051_PI426896222	AAGTATGTGAACTATGGGAGGAGCTCTCTGATGAAAACAGGTAAAATCTATGACCAGAAT
CUST_21001_PI426896222	CAGCAAGAATTCCAAGGCTTCTCATTTGTTAACCCTGAGTACTCTTACACAGATCACTGA
CUST_41366_PI426896222	TGTAAGAGGTCATGTCTTAAGCCAAACACAGTTGACCAACTAGTCTTCCTTGCACTCAAC
CUST_41052_PI426896222	AAAGAAAGAGGAGAACAAGAGCTTCACTGATGTGATGGAGTACAGTCCAGACACTGGGAA
CUST_21060_PI426896222	ATGAATTTGTCCAACCAAAGGAGCTGATGGAGGCAAGCAGTTCTGAAGGATCTGTTGATG
CUST_14085_PI426896222	TCCACGTCGCCGTTGTCGGTCTTTCTAGTGTTCCAAGAGTAGTTTTTCTCTCTCTTTTTT
CUST_22935_PI426896222	AGACTCTGCACAGAATTCAACAAATGGGAGTGGGAATTCAAAAAACAAATGTACACCTGG
CUST_27670_PI426896222	TACTACTTCTTCATCTGAAAGGGCTTTCAGTGCGAGTGGGAACATCGTCACATGTAAGAG
CUST_25225_PI426896222	GATCACACAAACTTCAACATTCAGCATCACAGTACTGAAGTGGACTCTATTAACCACACA
CUST_39928_PI426896222	TTTGTGTGGACAGTTTTTTAAATACATGCCATCTCTGTGATCCGTCAAAAACATTGTGTT
CUST_16862_PI426896222	CTTACCCCTGACTCTGATGTGCTGATGTACAACGACACAGTCATCACTGCAGACCTGTAA
CUST_28545_PI426896222	AACCTTCGAAGCGAGCTTCTTTGAACTCGTCCGTCATTTGGGAATAAAGCTGTTCGTCTT
CUST_25149_PI426896222	CAGGAAAATTCTGCGTAACTGAGGGCATGAAATATCTGAGGATGAGAGGGCATGAAGACA
CUST_6724_PI426896222	AAGATGCTGAGAAAAGCCAGAACTCTCCAACTAAAGAAGAGCAGACTCAAGCTGAGGGAA
CUST_28144_PI426896222	AAGGAGAAATCTTTTAGAACCCCAGAAGCAGTCTCAGCTTACCTGGATAACGTGCCTCAA
CUST_10109_PI426896222	CAAGCCCTCTCCATCACTCAGCATTAACTGCAGTGTGGTTACGTAAGCAAACTGCTAATT
CUST_5193_PI426896222	TGTACAATGCCAGGCTCAGTTCAGTTTCAAGAATGTGCTGTCACTGACAGAGAAGGTATC
CUST_13137_PI426896222	GTTAAAGGTGCAGTGTCAATCTTGCCAAAGATGGGCACATTTTGCATGTGCGAAATATAA
CUST_555_PI426896222	AGGTTGAAACTTTACGCTGAGACCCCAACTTAAAAGTGTCATAGCAGTAATCATTCAAAA
CUST_7071_PI426896222	ACTGACAATTCCCCTACAGTACACCTGTTTGGGTTTCAACGACTTTAAAGGACAATTCAG
CUST_24290_PI426896222	ACTGACAATTCCCCTACAGTACACCTGTTTGGGTTTCAACGACTTTAAAGGACAATTCAG
CUST_35393_PI426896222	TTCCATTTAGCGAAAAAACAAACAAACAAAAAGGGTGGGACTCGGTGGGAGGAGAGGGTT
CUST_40351_PI426896222	CAAAGATGTAAGTGTGATCGCTTCAAACTGAAGTCTTTCAAACTCACACGTAAAATCTTC
CUST_6825_PI426896222	ATGTCAATTGCATCCAGTTTGGTGGGTGAGGATGCCAAGACCAAGTTTCTAAGTAAGATG
CUST_7935_PI426896222	ATTCACCGAAGATGCATTGTATCTGGCATCATACGAGAGCGAAGGTCCTGAGAACAACAT
CUST_452_PI426896222	TTTCTAACAGGTAATCTTCTAACCCTTTCCAATTCTACCAAATAAAGCGTCTGTGCTTTG
CUST_5723_PI426896222	CTCTTCAGGGTCTCCCAGGAGTTTTAAAAGCATTGATTTCATTAATCTGAAAAAAGCATA
CUST_30073_PI426896222	TACAAACAAATGCCGGAGCCCAAATACACTCTCAGAAAGAAGCTCGTCCTCAAATTTGTA
CUST_23873_PI426896222	AAGAACGGACAGAAGGTCACAGAAGGATCCAGCACCAACACAGATATTTTTGTTCTTTGA
CUST_20878_PI426896222	ATTGAATCCATTGAGAAGGCAGTGGAGGACAAGATTGAGTGGTTGGAGTCCCACCAAAAT
CUST_5096_PI426896222	CCTGGCATATGTTAATTTTGCTCTGGTTGTGTACACGTTCTGTATACTGTTTCTCAATAA
CUST_31928_PI426896222	TCGAACTGGTTAACCTAAAACGTCTGAACCTGTCCAGGAATCCAATCCTCTGCATCTGTT
CUST_9238_PI426896222	CCTTTTGTAAATGAAACTCGTACTGTTGTAATGTTTCGTCTTTGGATGCCAAGCTGTGTA
CUST_13493_PI426896222	AGCTGTGACAACTGTGTAGAGCTTGACATAATGTCCATAGCGGGCATGGTCATAGGAAAT
CUST_3491_PI426896222	GTGGGGAAAGACGAATAAAACTATGGTGCATTTTGTTACAACTGTTTCACATTTATGGAA
CUST_35899_PI426896222	CATCCGATTCACATCATCTCTTTGTTTGAAGCAGTTTCATAGGATTACAAGAGGTGTAAA
CUST_800_PI426896222	TTAGAATGGCTTTTGTAGAGTCGCTGTGCTCATGAGCATGATCAGTTCATTTGCAGCAAA
CUST_30605_PI426896222	ATAAAGAGAAGAAGAACGGCAACCACTTGCTCCAGCTGTGGAAAGAGTGTCTGCAAAGAT
CUST_8023_PI426896222	CATGAGGTCATCTGGTGGAGCATAATGATGCTAATGATGCTTTTATATTTGTTTTGGTAA
CUST_23235_PI426896222	ACTTACTCACAGAACAACAGCTCCAAATAAAGACTTTTCTTCAACTGACTCTGTTATGTA
CUST_13679_PI426896222	GCCTCATTTTTTAATCTTTGTGCTAAGCTAGGCAACAGTATTTGAGAGCTAGCTGTTTAT
CUST_24221_PI426896222	TTACAGTGTGTGAAGTTTGAAGTTTGGAAATGTCTGATAAGGGAAGAGGAGGACCAAGTG
CUST_5738_PI426896222	TTGAGTGGTACATAACAGGAAACATTTCATCGTCATCCTGGTTTGACCACGTCAGGGAGT
CUST_25471_PI426896222	ACTCTTCACTGTCAGTGTTTCAAGGCATGTCAAGGCATCAAATAAAAACAGAAAAACATC
CUST_1524_PI426896222	GGAACAGTTTTGTGTATTAAGGACTGTTAGTTTGGAAAGTTGGATCTGTAAAATGGTCCC
CUST_2144_PI426896222	ATCATCAAGGAAATAGAAATATTCCCAGTGGTGAGGGATGACGCTCCAATAGAGAAGAAG
CUST_9317_PI426896222	GGCCACCTACAATAAGCTCTTCTTTAGGTTGGCTTTTGGACATTTTGAATTAAATGGATC
CUST_1135_PI426896222	CGTGTCTGTTAGCTGGAACGCATCAAAATGAAAGTGAATCCGTGAAATGTTGAATTGTCG
CUST_26098_PI426896222	AAGGATTCGGCTCATGAAGAACTTGACCAGAAGAATTCGGCTCATGAATAATATCACTTG

Ensembl ID ENSONIG000000…	NCBI accession No.	Tilapia LOC no.	Gene symbol	Relevance to SD References
18017	XM_003448079	LOC100707727	*tspan8*	[62]
8406	XM_005464738	LOC100712435	*fcgrt*	
4781	DQ257619	LOC100707206	*amh*	[21, 7]
20797	XM_005466025 & XM_005462477	LOC102082816 & LOC100709289	*urgcp & gvinp1*	
9031			*ctnnd2b*	S1
12502 or 08413	AF202719		*g3s7a7*	
3316	XM_005460744	LOC100703205	*prkcb*	[65]
12573, 07813, 05031	XM_005475764	LOC102082186	*zbed3*	S2, [63]
276	XM_005466392	LOC102077119	*urgcp*	
01911[N]	XM_003456007	LOC100711688	*cyp4f3*	[64]
13440	XM_003449942	LOC100693709	*wasp1*	S3
19899	XM_005462508	LOC102081652	*urgcp*	
12573	XM_005475764	LOC102082186	*zbed3*	S2, [63]
4507	XM_005465472	LOC100698188	*cdh1*	
20219	XM_005466007	LOC100697657	*gpr144*	
8814	XM_003452343	LOC100691473	*ror1*	
6564	XM_003455332	LOC100703378	*fam131c*	
	XM_003443600	LOC100692611	*pkia*	
6534	XM_003446414	LOC100706747	*eci1*	
980	XM_005468875	LOC100699105	*itgb3a*	
1216	XM_005462079	LOC102083022	*timm50*	
10126	XM_005476196	LOC102077712	*fndc7*	
	XR_266011	LOC102078532	*loc102078532*	
2303	XM_003444133	LOC100698476	*nkain1*	
14484	XM_003443593	LOC100690454	*kiaa1468*	
14048	XM_003444539	LOC100696614	*rasgef1b*	
9474	XM_003438088	LOC100693488	*higd1a*	
9725	XM_003442602	LOC100710565	*sult3*	
15679	XM_003457459	LOC100702139	*casp8*	
18452	XM_005463621	LOC100693481	*orni-dba*	
12430	XM_003455831	LOC100705893	*hspa5*	
19699	XM_005448344	LOC100698718	*chadl/ socs7*	[69]
03083[N]	XM_003445709	LOC100695831	*atp5g3*	
16686	XM_005454329	LOC102080392	*cd3d*	
551	XM_003440733	LOC100703324	*loc100703324*	
6329	XM_005461525	LOC100707490	*loc100707490*	
10523	XM_003437773	LOC100697473	*mrpl34*	
	XM_005467716	LOC100692887	*selh*	
20048[N]	XM_003458001	LOC100697999	*psmb8*	
19128	XM_003449410	LOC100707903	*ankrd22*	
07160[N]	XR_267867	LOC100690239	*loc100690239*	
9725	XM_003442602	LOC100710565	*sult3st1*	
11729	XM_003459954	LOC100695718	*rtn4ip1*	[67]
16642	XM_005448435	LOC100710376	*zp3*	[70]
6350	XM_003439269	LOC100696589	*impa1*	[68]
3386	XM_005466820	LOC102079786	*gpa33*	
1577	XM_005473633	LOC100534432	*cr/20ß-hsd*	[66]

Relevance to apoptosis / immune response/ TGF-β		Notes
Description	References	
	S4	
Encodes for receptor that binds the Fc region of monomeric immunoglobulin G that is involved in macrophage/fibroblast-mediated apoptosis	S5	
Key player in cell proliferation, differentiation and apoptosis	[71]	
MHC CLASS I Family (ENSFM00500000269621)	S6	Ortolog of FCGRT
SMAD-dependent TGF-β signaling is modulated by PRKCB		
	S7	
Humal Orthologue SMAD3 in part of TGF-β pathway	[69]	
	S8	
TGF-β	S9	
Humal Orthologue SENP1	S10	
Human p27 via the Ral-GEF pathway and disrupts TGF-β-mediated Smad nuclear translocation	S11, S12	
	S13	
Conserved function of caspase-8 in apoptosis during bony fish evolution	S14	
MHC CLASS II Alpha (ENSFM00350000105424)	S15	
A master regulator of the anti-apoptotic unfolded protein response signalling network	S16	
Suppressors of cytokine signaling (SOCS) in T cell differentiation, maturation, and function	S17	
	S18	
SelH protects neurons against UVB-induced damage by inhibiting apoptotic cell death pathways	S19	
The immunoproteasome, PSBB8, appears to be a key link between inflammatory factors and the control of vascular cell apoptosis	S20	
Can modulate anti-apoptotic activity		
Humal Orthologue C4BPA	S21	
Member of the immunoglobulin superfamily,	S22	

Array probe	Microarray probe name deposited in NCBI's Gene Expression Omnibus accession No. GSE50974
Probe sequence	Microarray probe sequence deposited in NCBI's Gene Expression Omnibus accession No. GSE50974
Ensembl ID	The informative digits of ENSEMBL gene ID obtained following a probe sequence BLAT search
NCBI accession No.	NCBI gene ID obtained following a probe sequence BLAST search
Tilapia LOC no.	Tilapia protein no. obtained from NCBI Gene database
Gene symbol	Gene symbol used in current work, based on NCBI, ENSEMBL or tilapia LOC no.
Linkage group	Probe location on tilapia linkage group (LG) obtained following probe sequence BLAST search (http://cichlid.umd.edu/blast/blast.html)
Physical location	Probe location on tilapia genome obtained following probe sequence BLAST search (http://cichlid.umd.edu/blast/blast.html)
Description	Gene full name
Human protein	Obtained by BLAST against human proteins using the amino-acid sequence of the NCBI accession
Human protein symbol	Human protein symbol
FC	Fold change of female to male expression Least Squares (LS) means
P value	Nominal significance value of female to male expression LS means
FDR	False discovery rate as a function of nominal significance values
Relevance to SD	Description and references (in the manuscript [] and in the bottom of this Table S1-S2)
Relevance to apoptosis / immune response/ TGF-β	Description and references (in the manuscript [] and in the bottom of this Table S3-S20)
Notes	Additional information
N	Gene annotation was based on location of nearest gene (Ensembl probe sequence by BLAT search)

S1. Sun F, Liu S, Gao X, Jiang Y, Perera D, et al. (2013) Male-Biased Genes in Catfish as Revealed by RNA-Seq Analysis of the Testis Transcriptome. Plos One 8: e68452.

S2. Chen Y, Webster TJ (2009) Increased osteoblast functions in the presence of *BMP-7* short peptides for nanostructured biomaterial applications. Journal of Biomedical Materials Research Part A 91A: 296-304.

S3. Ochs HD, Thrasher AJ: The Wiskott-Aldrich syndrome. Journal of Allergy and Clinical Immunology 2006, 117:725-738.

S4. Yue S, Mu W, Zöller M (2013) *Tspan8* and *CD151* promote metastasis by distinct mechanisms. European Journal of Cancer.

S5. Attisano L, Labbé E (2004) TGFβ and Wnt pathway cross-talk. Cancer and Metastasis Reviews 23: 53-61.

S6. Schaible UE, Winau F, Sieling PA, Fischer K, Collins HL, et al. (2003) Apoptosis facilitates antigen presentation to T lymphocytes through *MHC-I* and *CD1* in tuberculosis. Nature medicine 9: 1039-1046.

S7. Sato R, Iiizumi S, Kim E-S, Honda F, Lee S-K, et al. (2012) Impaired cell adhesion, apoptosis, and signaling in *WASP* gene-disrupted Nalm-6 pre-B cells and recovery of cell adhesion using a transducible form of WASp. International journal of hematology 95: 299-310.

S8. Daneshmanesh A, Hojjat-Farsangi M, Khan A, Jeddi-Tehrani M, Akhondi M, et al. (2012) Monoclonal antibodies against *ROR1* induce apoptosis of chronic lymphocytic leukemia (CLL) cells. Leukemia 26: 1348-1355.

S9. Heino J, Ignotz RA, Hemler ME, Crouse C, Massague J (1989) Regulation of cell adhesion receptors by transforming growth factor-beta. Concomitant regulation of integrins that share a common beta 1 subunit. Journal of Biological Chemistry 264: 380-388.

S10. Li X, Luo Y, Yu L, Lin Y, Luo D, et al. (2008) *SENP1* mediates TNF-induced desumoylation and cytoplasmic translocation of *HIPK1* to enhance ASK1-dependent apoptosis. Cell Death & Differentiation 15: 739-750.

S11. Kfir S, Ehrlich M, Goldshmid A, Liu X, Kloog Y, et al. (2005) Pathway- and Expression Level-Dependent Effects of Oncogenic N-Ras: p27Kip1 Mislocalization by the Ral-GEF Pathway and Erk-Mediated Interference with Smad Signaling. Molecular and Cellular Biology 25: 8239-8250.

S12. Yaman E, Gasper R, Koerner C, Wittinghofer A, Tazebay UH (2009) *RasGEF1A* and *RasGEF1B* are guanine nucleotide exchange factors that discriminate between Rap GTP binding proteins and mediate Rap2 specific nucleotide exchange. FEBS Journal 276: 4607-4616.

S13. Bedó G, Vargas M, Ferreiro M, Chalar C, Agrati D (2005) Characterization of hypoxia induced gene 1: expression during rat central nervous system maturation and evidence of antisense RNA expression. International journal of developmental biology 49: 431.

S14. Sakata S-I, Yan Y, Satou Y, Momoi A, Ngo-Hazelett P, et al. (2007) Conserved function of *caspase-8* in apoptosis during bony fish evolution. Gene 396: 134-148.

S15. Nag B, Kendrick T, Arimilli S, Yu SCT, Sriram S (1996) Soluble MHC II–Peptide Complexes Induce Antigen-Specific Apoptosis in T Cells. Cellular immunology 170: 25-33.

S16. Uckun FM, Qazi S, Ozer Z, Garner AL, Pitt J, et al. (2011) Inducing apoptosis in chemotherapy-resistant B-lineage acute lymphoblastic leukaemia cells by targeting *HSPA5*, a master regulator of the anti-apoptotic unfolded protein response signalling network. British Journal of Haematology 153: 741-752.

S17. Palmer DC, Restifo NP (2009) Suppressors of cytokine signaling *(SOCS)* in T cell differentiation, maturation, and function. Trends in Immunology 30: 592-602.

S18. Takada H, Nomura A, Roifman CM, Hara T (2005) Severe combined immunodeficiency caused by a splicing abnormality of the *CD3δ* gene. European journal of pediatrics 164: 311-314.

S19. Mendelev N, Witherspoon S, Li PA (2009) Overexpression of human *selenoprotein H* in neuronal cells ameliorates ultraviolet irradiation-induced damage by modulating cell signaling pathways. Experimental neurology 220: 328-334.

S20. Huelsken J, Birchmeier W (2001) New aspects of Wnt signaling pathways in higher vertebrates. Current Opinion in Genetics & Development 11: 547-553.

S21. Rezende SM, Simmonds RE, Lane DA (2004) Coagulation, inflammation, and apoptosis: different roles for protein S and the protein S–C4b binding protein complex. Blood 103: 1192-1201.

S22. Heath JK, White SJ, Johnstone CN, Catimel B, Simpson RJ, et al. (1997) The human A33 antigen is a transmembrane glycoprotein and a novel member of the immunoglobulin superfamily. Proceedings of the National Academy of Sciences 94: 469-474.

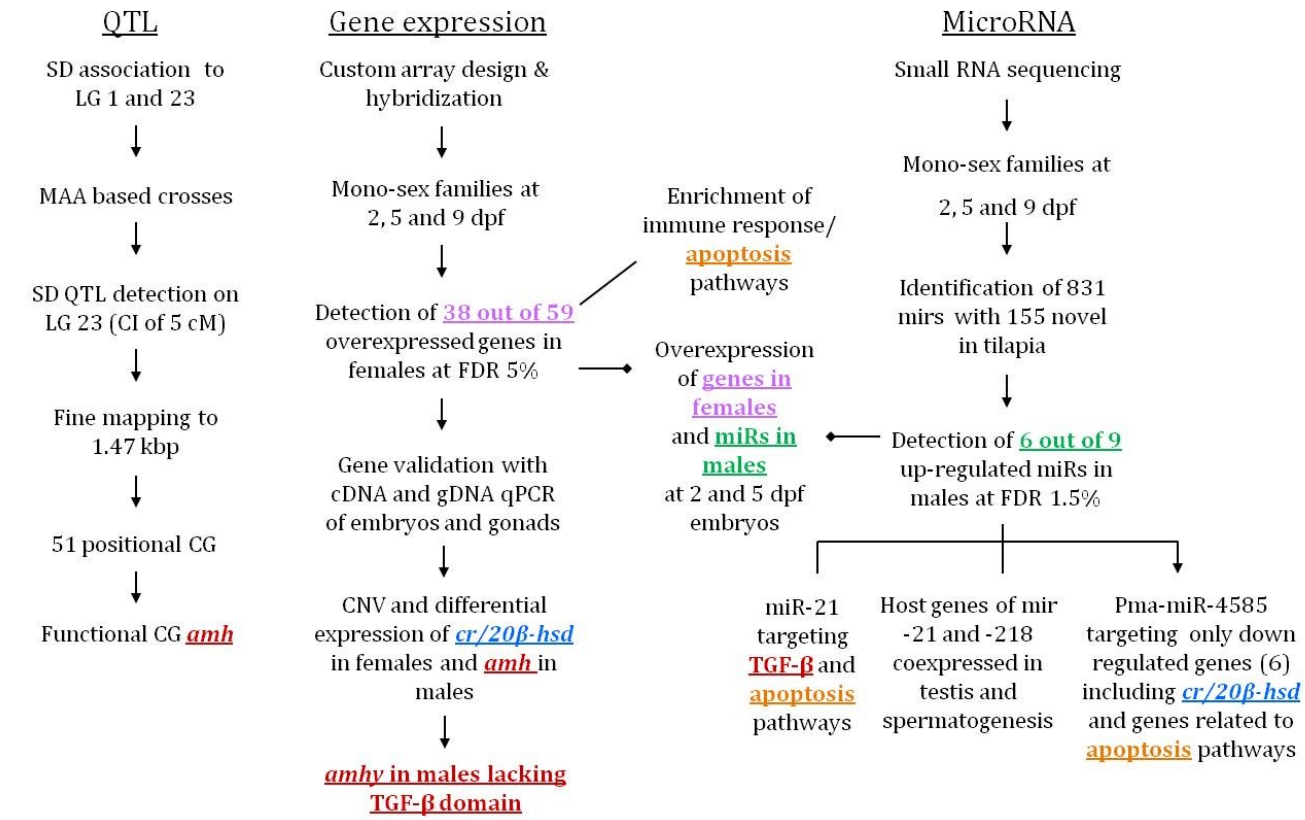
Overall discussion

Detection of SD genes in tilapia has both scientific and commercial importance. In this study, several approaches were employed to elucidate SD in tilapia including SD QTL identification, QTL fine mapping, sexually differential expression patterns of genes and miRNAs, and identification of CNV. Genes and miRNAs involved in SD have been searched at the early developing embryo when the SD mechanism is being initiated. The experimental workflow of the study and the main results are presented in Figure 1. QTL on LG23 identification and fine mapping is displayed on the left axis, gene expression analysis is presented in the middle with the miRNAs analysis on the right. Results and conclusions integrated from different analyses are presented by a common color.

QTL for Sex Determination

QTLs for SD were detected on LG1 (microsatellite marker *BYL002* from Kocher's cichlid website genomic browser: scaffold_17:190588..191029; P = 0.01) and LG23 (microsatellite marker *UNH898*; P = 8.6 x 10^{-5}). One of the *UNH898* alleles was termed MAA (Male Associated Allele) since it was mainly carried by males. The genotypes of sex reversed individuals were confirmed based on presence or absence of MAA through different crosses followed by progeny testing. Integrated mapping analysis of this locus based on three families showed a maximum F-value of 42 on LG23 at 18 cM, with a 95% confidence interval of 16–21 cM (Eshel *et al*., 2011). In a further fine-mapping study we genotyped our segregating family of 90 individuals for additional microsatellite genetic markers. The sex associated region was localized to scaffold 101 between markers *GM597* and *ARO124* from 990,577 to 2,468,000 bp. Twelve adjacent markers found in this region were homozygous in females and either homozygous for the alternative allele or heterozygous in males. Markers flanking the critical sex region were heterozygous in two females thus localizing the QTL on LG23 into a 1.47 Mbp region (Eshel *et al*., 2012). This genomic region harbored 51 positional candidate genes including a single functional candidate gene (*amh*). The ability to further narrow down this interval below a cM unit requires sampling of thousands of fish to allow detection of recombinants for the QTL.

Figure 1. Experimental workflow for detection of SD QTL, and differentially expressed genes and miRNAs between genders.



Legend - SD: Sex determination, LG: linkage group, MAA: male associated allele, QTL: quantitative trait loci, CG: candidate gene, mir: miRNA precursor, miR: expressed 3p/5p mature sequence from miRNA precursors, *amh*: Anti-Müllerian hormone, *cr/20β-hsd*: carbonyl reductase 20 beta-hydroxysteroid dehydrogenase, TGF-β: transforming growth factor beta, FDR: false discovery rate, CNV: copy number variation. The implicated conclusions integrating the three streams of data are presented by a common color.

→Experimental flow; —◆ Implicated from the results; —|Implicated from the literature.

Integration of our results and other publications addressing SD-associated markers raised the question: How SD may be almost fully explained by QTL from different genomic regions in different families? For example, SD-associated markers *UNH995* and *UNH104* on LG1 explained the sex of 95% of individuals in two *O. niloticus* families (Lee *et al*., 2003). A third family from the same population showed no evidence of linkage for this region with phenotypic sex, which may indicate a non-segregating QTL (homozygous). Studies in *O. aureus* and F_2 family derived from *O. aureus* x *O. niloticus* cross identified a QTL for SD mapped to LG3. However, markers on LG1 in *O. aureus* also showed a strong association with sex, indicating segregation of a male-determining allele in both of these regions (Lee *et al*., 2004). Cnaani *et al*. (2004) detected five markers with association to sex in an F_2 population derived from an interspecific tilapia hybrid (*O. mossambicus* x *O. aureus*). The most significant markers, *UNH868* and *UNH925*, from LG 1 and 3, respectively, explained 34% of the sex variation. Lee *et al*. (2005) and Lee and Kocher (2007) detected SD QTL on LG1 in *O. niloticus* and fine mapped it to a 2.6 cM region, but failed to identify the causative gene. Shirak *et al*. (2002) suggested that *UNH216* and *UNH231* on LG 23 and 6, respectively, are linked to sex ratio distortion genes, and that *UNH159* on LG2, may be linked to a modifier of these genes in a full-sib family of 222 progeny from the fourth generation of a meiogynogenetic tilapia line (*O. aureus*). Eshel *et al*. (2011) found SD QTL on LG23 ($p=8.6x10^{-5}$), explaining 97.4% of the sex variation and additional QTL on LG1 ($p=0.01$). Recently, Palaiokostas *et al*. (2013) identified a major sex-determining region on LG1, explaining 96% of the phenotypic sex variance based on restriction associated DNA (RAD) sequencing and mapping. Additionally, numerous studies found that elevated temperatures can alter individuals' SD indicating that additional genetic and/or environmental factors regulate SD (Reviewed by Baroiller *et al*., 2009).

Sex and sex ratio can be explained by a dominant gene (*SRY* in human), gene dosage (*Drosophila*), environmental influence (Alligator), or by the 'threshold dichotomy' theory (Mittwoch, 2006), that applies to a trait with contrasting phenotypes originating from multiple genes with quantitative effects. Furthermore, *O. niloticus* and *O. aureus* have different sex chromosome systems and their ability to mate and produce fertile hybrids further complicates the elucidation of the SD system. The complexity of SD and the limitations of QTL mapping (reviewed by Weller and Ron, 2011) complicate the identification of the causative genes for SD. The candidate gene approach is also rather weak since it relies on knowledge on the function of genes in the biological pathways of SD that are apparently poorly understood in fish. Hence, gene expression approach may be used for the identification of genes that are involved in SD without *a priori* prediction or interpretation.

Differential Expression of Genes Between Genders

Differentially expressed genes between gonads in tilapia are well documented in the literature (Baroiller *et al*., 1995, 2009; Poonlaphdecha *et al*., 2013). Since sexually differential expression patterns of genes were detected in the bi-potential gonad at 10 dpf (Ijiri *et al*., 2008), we assumed that genes initiating the mechanism of SD are expressed earlier. In a preliminary study we detected differentially expressed SD candidate genes e.g., *elavl1*, *amh*, *lhx9*, *sox9*, *foxl2* and *sox14* at early developmental stages (≤ 9 dpf) using qPCR (Eshel *et al*., 2012). These results confirmed that genes initiating the SD cascade are expressed before gonad differentiation and that the time window of 2 to 9 dpf is adequate for the transcriptome-wide microarray experiment.

An Agilent eArray with 43,803 probes was constructed representing the whole transcriptome including 51 candidate genes from LG23 and genes relevant to SD inferred from other organisms, to detect genome-wide differentially expressed genes between genders. The main weakness of microarray is that only represented genes may be detected. Since the tilapia genome is only partially annotated, some genes were not represented. Additionally, in case of alternative splicing, where only one transcript has a role in SD, it may not be detected as a differentially expressed gene. Thus, by pursuing the microarray experiment we found 59 genes differentially expressed between genders. Genes that were highly expressed in males were *amh*, *tspan8, zbed3, wasp1, cyp4f3* and *prkcb*, and the most significantly overexpressed genes in females were *cr/20β-hsd*, *rtn4ip1*, *impa1, socs7* and *zp3*. All of these genes were associated to either sex related pathways (fertility, mating etc.), TGF-β pathway or apoptosis/immune response (highlighted in blue, red and orange, Figure 1). Both immune response and TGF-β pathway are related to apoptosis (Schaible *et al*., 2003; KEGG database). Functional annotation clustering indicated a significant enriched cluster of the immune response containing four genes: *psmb8*, *fcgrt*, *gas7a7* and *zp3*. This is in accordance with previous studies, suggesting that evolutionary conserved genes in the immune system and apoptotic cell death processes may also play a role in this early stage of differentiation and SD (Opferman, 2008). Twelve of the 59 differentially expressed genes were characterized experimentally or bioinformatically with CNV. CNV and dosage sensitivity have been hypothesized as evolutionary conserved factors of SD and SD plasticity among related species (Schartl, 2004; Volff *et al*., 2007; Hattori *et al.,* 2013). The first identified three SD genes were *SRY*, *Dmw* and *dmy* in human, birds and medaka, respectively. *SRY* has been suggested to evolve from an extra copy of *SOX3* (Foster and Graves, 1994; Graves, 1998), and the latter two evolved from *DMRT* gene duplication (Nanda *et al*., 2000, 2002; Smith *et al*., 2009). Interestingly, *DMRT* sex specific duplications were found both in XX/XY as well as

ZZ/WZ sex chromosomal systems (Nanda *et al*., 2002; Yoshimoto *et al*. 2008). Hong *et al*. (2007) reviewed *DMRT* genes' function in various species and reported their role in apoptotic pathways. Our findings, supported by various studies, indicate that CNV is a common feature of genes participating in SD and may be the alternative genomic structure to sex chromosome systems in fish.

One of the genes with CNV was *amh* that in our earlier study was mapped to SD QTL on LG23 (Eshel *et al*., 2011, 2012) (highlighted in red, Figure 1). *amh* is a member of the TGF-β that is a key player in cell proliferation, differentiation and apoptosis (Kubiczkova *et al*., 2012). It was found to be expressed in both ovary and testis but is significantly overexpressed in male embryos. *AMH* has a role in female fertility and its plasma level reflect ovarian reservoir in human (reviewed by Grynnerup *et al*., 2012). Fish have *amh* even though they lack the Müllerian duct. Although *amh* function is still unknown in fish, it has a major role in SD and was suggested as a master regulator of SD (Shirak *et al*., 2006). In this work we identified an *amh* unique copy, denoted *amhy,* differing from the original gene by a 233 bp deletion of the TGF-β domain. *amhy* was detected only in the male genome, and expressed in testis, whereas, *amh* was present in both genders' genome and expressed in ovary as well as testis (highlighted in red, Figure 1). Recent works identified *amh* and *gsdf* (member of TGF-β family) male specific copies in *Odontesthes hatchery* and *Oryzias luzonensis*, respectively (Hattori *et al*., 2012; Myosho *et al*., 2012). The injection of an *amh* antisense morpholino (MO) to XY embryos resulted in ovary development in *Odontesthes hatcheri*and (Hattori *et al*., 2012), and the presence of a genomic fragment that included *gsdfy* converts XX individuals into fertile XX males in *Oryzias luzonensis* (Myosho *et al*., 2012). These results propose that members of TGF-β family have a critical role in SD cascade in fish.

The presence of additional *amh* copy with major variation raises the question of how it affects SD. *Amh* functions primarily through *amhr*2 (Mishina *et al*., 1999). Upon binding to *amh*, *amhr2* recruits and phosphorylates a type I receptor(s) that then transduces signals by phosphorylating Smad proteins which in turn regulate transcription of downstream genes in mammals (Belville *et al*., 2009). Loss-of-function of *amhr2* in male mouse leads to a partial hermaphrodite having a uterus and an oviduct together with the testis (Jamin *et al*., 2003). However, in medaka, a homozygous mutation in exon 9 of *amhr2* results in complete sex reversal in half of the genetic males (Morinaga *et al*., 2007) and in fugu (*Tiger pufferfish* and *Takifugu rubripes*) missense SNP in this gene was found to be associated with SD (Kamiya *et al*., 2012). These reports suggest evolutionally conserved functions of the *amh*/*amhr*2 signaling network, such as ovarian folliculogenesis and gonadal steroidogenesis.

Interestingly, we found that *cr/20βhsd* (highlighted in blue, Figure 1) was overexpressed in female embryos and ovary and showed high genomic copy number in females implying its main role in female development. This gene is known to be part of the oxidoreductase pathway for oocyte maturation preceding the enzymatic activity of *cyp19* (*Cytochrome P450 aromatase*) (Senthilkumaran *et al*., 2004). Kwon *et al*. (2001) found initiation of expression of both brain and ovarian aromatase mRNA between 3 and 4 dpf both in males and females, but no significant sexual differences were detected up to 11 dpf. Rodriguez-Mari *et al*. (2010) outlined a SD model in zebrafish where apoptosis, *cyp19a1a* and *amh* were proposed as the main factors controlling SD.

Identification of Differentially Expressed microRNA Between Genders

The emerging significance of miRNAs in developmental processes has attracted research in various organisms (Wienholds *et al*., 2005; Bannister *et al*., 2009; McFarlane and Wilhelm, 2010; Morgan and Bale, 2012). The ability of miRNAs to regulate large number of genes makes them an important factor for consideration in the SD process. Tilapia miRNAs are not publicly available although recently Huang *et al*. (2012) performed next-generation sequencing to define the first miRNA transcriptome consisting of 184 miRNAs in skeletal muscle of Nile tilapia, and Yan *et al*. (2012) identified 25 conserved miRNAs in tilapia skeletal muscle using small RNA cloning. By sequencing of small RNA from the same biological samples that were hybridized to the microarray, 831 miRNAs were found in tilapia embryos at 2, 5 or 9 dpf, of which 155 were novel. Our findings of nine miRNAs that are differentially expressed between genders from 2 dpf illustrate their role in the early developing embryo. Four differentially expressed miRNAs had annotations. For pma-miRNA-4585, an overexpressed miRNA in males, all six predicted target genes, including *CR/20β-HSD* that is known to activate oocyte maturation (highlighted in blue, Figure 1), as well as *RTN4IP1* and *PSMB8* that function in apoptosis pathway (highlighted in orange, Figure 1), were down-regulated in males. This miRNA showed significantly perfect inverse expression correlation with its targeted genes, in accordance with the expected inhibition of mRNA translation of target genes (Plasterk, 2006; Guo *et al.*, 2010). Two of the four miRNAs with annotations were conserved and thus allowed genomic analysis of their host genes in a variety of vertebrate species. The host genes of miRNA-218 and miRNA-21, which were overexpressed in males, were *DRG1* and *TUBD1*, respectively. Both genes function in male developmental processes related to sperm and testis (Smrzka *et al*., 2000). Moreover, Papagiannakopoulos *et al*. (2008) reported that miRNA-21 targets several genes in the TGF-β and

apoptosis pathways (highlighted in red and orange, respectively, Figure 1). These results support the differential expression of these miRs, similar to the identification of various differentially expressed miRNAs between genders in chicken embryos (Bannister *et al*. 2009).

Interestingly, 38 out of the 59 differentially expressed genes by genders were overexpressed in females (highlighted in purple, Figure 1), and that six out of the nine miRNAs were overexpressed in males at the same period of early embryonic development (highlighted in green, Figure 1). These results indicate an early onset of genes in the female determination, which may be the default sex determiner in XX/XY chromosomes system, unless they are down-regulated by miRNAs thus initiating the male determining pathways. For example, in *Odontesthes hatchery*, histological sex differentiation of the gonads showed that the ovary differentiated at 3-4 weeks after hatching, as compared to 5-6 weeks in testis (Hattori *et al*., 2012). The presence of oocytes appears to be important for sex determination for zebrafish and medaka. In zebrafish all embryos start to develop as females, and in medaka only XX females start oogenesis while XY males suppress oogenesis and all germ cells remain undifferentiated (Siegfried and Nusslein-Volhard, 2008; Saito and Tanaka, 2009). Furthermore, the number of developing oocytes is a key feature that signals the ovary fate of undifferentiated gonads (Rodrıguez-Mari *et al*., 2010).

Proposed Future Work on Various Aspects of Sex Determination

SD is influenced by several genetic and environmental factors, with variation between families and related strains. So far the search for QTL for SD has located large regions in different linkage groups that were inconsistent between strains and crosses of tilapia (Lee *et al*., 2003, 2004; Cnaani *et al*., 2004, 2008; Eshel *et al*., 2011; Palaiokostas *et al*., 2013). Fine mapped regions of these QTL still contained dozens of genes (Lee and Kocher, 2007; Eshel *et al*., 2012). Genome-wide association studies (GWAS), in which several hundred thousand to a million single nucleotide polymorphisms (SNPs) are assayed for unrelated individuals, represent a powerful tool for investigating the genetic basis of complex traits (Manolio *et al*., 2009). However, this approach is dependent on commercial development and manufacture of a SNP chip which is not currently available for tilapia. The limitation of GWAS is that genes with CNV are excluded from statistical analysis since SNPs deviating from Hardy-Weinberg equilibrium cannot be distinguished from technical errors. Enrichment of CNV in tilapia genome and particularly in SD genes may interfere with the application of GWAS once a SNP chip is available.

In order to study the mechanism of SD, transgenesis and siRNA technologies may be used. Transgenesis is an excellent tool to evaluate specific gene's influence on whole embryo, which was recently demonstrated for Nile tilapia (Fujimura and Kocher, 2011; Golan and Levavi-Sivan,

2013). For example, the role of *amhy* as our proposed key regulator of SD in Nile tilapia can be verified by insertion of this gene to female embryos lacking this gene, thus producing males. Additional tool is gene silencing using siRNA technology, as applied in the giant freshwater prawn (Ventura *et al*., 2009).

Our work clearly demonstrates that miRNA have a role in SD that is still unknown in vertebrates. Their function and impact on individual's phenotype, specifically its gender, can be evaluated by silencing miRNAs *in vivo*, as applied in Nile tilapia (Yan *et al*., 2012). Identification of miRNAs' target genes can be analyzed by cross-linking immunoprecipitation (HITS-CLIP, Thomson *et al*., 2011). Environmental factors are known to alter sex ratios (D'cotta *et al*., 2001; Devlin and Nagahama, 2002). Working with gene expression microarray or total RNA sequencing and small RNA sequencing in monosex families raised in different temperatures may discover the affected genes and miRNAs.

Summary

This study reports the fine-mapping of QTL on LG23, the discovery of sexually-dimorphic expression patterns of genes and miRNAs, and genes enriched for CNV. This is the first report of male-specific *amh* duplication and detection of SD pathways that are functional at 2 to 9 dpf tilapia embryos. Thus, the sequence of events leading to SD in tilapia is apparently initiated soon after fertilization. Systems biology techniques were used to derive and connect information on QTLs, genes and miRNAs. The implicated conclusions integrating the three streams of data are presented by a common color in Figure 1. *amh*, *cr/20β-hsd* and 22 genes involved in apoptosis were shown to be differentially expressed and constitute targets for sexually-dimorphic miRNA. *amh*, *cr/20β-hsd* and apoptosis are the major genes and mechanism (highlighted in red, blue and orange, respectively, Figure 1), that determine sex, similar to the proposed zebrafish SD model (Rodrıguez-Mari *et al*., 2010).

References

Anderson JL, Marí AR, Braasch I, Amores A, Hohenlohe P, et al. (2012) Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. PloS one 7: e40701.

Ayling L-J, Griffin D (2003) The evolution of sex chromosomes. Cytogenetic and Genome Research 99: 125-140.

Azuma T, Takeda K, Doi T, Muto K, Akutsu M, et al. (2004) The influence of temperature on sex determination in sockeye salmon *Oncorhynchus nerka*. Aquaculture 234: 461-473.

Bannister SC, Tizard MLV, Doran TJ, Sinclair AH, Smith CA (2009) Sexually dimorphic MicroRNA expression during chicken embryonic gonadal development. Biology of Reproduction 81: 165-176.

Baroiller JF, Chourrout D, Fostier A, Jalabert B (1995) Temperature and sex-chromosomes govern sex-ratios of the mouthbrooding cichlid fish *Oreochromis niloticus*. Journal of Experimental Zoology 273: 216-223.

Baroiller JF, D'Cotta H, Bezault E, Wessels S, Hoerstgen-Schwark G (2009) Tilapia sex determination: Where temperature and genetics meet. Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology 153: 30-38.

Baroiller JF, D'Cotta H, Saillant E (2009) Environmental effects on fish sex determination and differentiation. Sexual Development 3: 118-135.

Barske LA, Capel B (2008) Blurring the edges in vertebrate sex determination. Current Opinion in Genetics & Development 18: 499-505.

Beardmore JA, Mair GC, Lewis RI (2001) Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. Aquaculture 197: 283-301.

Bellott DWS, H. Pyntikova, T. Mardis, E. R. Graves, T. Kremitzki, C. Brown, L. G. Rozen, S. Warren, W. C. Wilson, R. K. Page, D. C. (2010) Convergent evolution of chicken Z and human X chromosomes by expansion and gene acquisition. Nature 466: 612-U613.

Belville C, Marécha JD, Pennetier S, Carmillo P, Masgrau L, et al. (2009) Natural mutations of the anti-Müllerian hormone type II receptor found in persistent Müllerian duct syndrome affect ligand binding, signal transduction and cellular transport. Hum Mol Genet 18: 3002–3013.

Berezikov E (2011) Evolution of microRNA diversity and regulation in animals. Nature Reviews Genetics 12: 846-860.

Birk OS, Casiano DE, Wassif CA, Cogliati T, Zhao LP, et al. (2000) The *LIM* homeobox gene *Lhx9* is essential for mouse gonad formation. Nature 403: 909-913.

Campos-Ramos R, Harvey SC, Masabanda JS, Carrasco LAP, Griffin DK, et al. (2001) Identification of putative sex chromosomes in the blue tilapia, *Oreochromis aureus*, through synaptonemal complex and FISH analysis. Genetica 111: 143-153.

Carrasco LAP, Penman DJ, Bromage N (1999) Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. Aquaculture 173: 207-218.

Charlesworth D, Charlesworth B (2005) Sex chromosomes: evolution of the weird and wonderful. Current Biology 15: R129-R131.

Charlesworth D, Charlesworth B, Marais G (2005) Steps in the evolution of heteromorphic sex chromosomes. Heredity 95: 118-128.

Charlesworth D, Mank JE (2010) The Birds and the bees and the flowers and the trees: lessons from genetic mapping of sex determination in plants and animals. Genetics 186: 9-31.

Chen T-R, Reisman H (1970) A comparative chromosome study of the North American species of sticklebacks (Teleostei: *Gasterosteidae*). Cytogenetic and Genome Research 9: 321-332.

Cnaani A, Lee BY, Zilberman N, Ozouf-Costaz C, Hulata G, et al. (2008) Genetics of sex determination in tilapiine species. Sexual Development 2: 43-54.

Cnaani A, Levavi-Sivan B (2009) Sexual development in fish, practical applications for aquaculture. Sexual Development 3: 164-175.

Cnaani A, Tinman S, Avidar Y, Ron M, Hulata G (2004) Comparative study of biochemical parameters in response to stress in *Oreochromis aureus*, *O. mossambicus* and two strains of *O. niloticus*. Aquaculture Research 35: 1434-1440.

Cnaani A, Zilberman N, Tinman S, Hulata G, Ron M (2004) Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. Molecular Genetics and Genomics 272: 162-172.

Crosetti D, Sola L, Brunner P, Cataudella S. (1988) Cytogenetical characterization of *Oreochromis niloticus*, *O. mossambicus* and their hybrid. In the second symposium on tilapia in aquaculture. ICLARM. Department of fisheries, Bangkok, Thailand and international center for living aquatic resources management, Manila, Filipinas. pp. 143-151.

Davey AJ, Jellyman DJ (2005) Sex determination in freshwater eels and management options for manipulation of sex. Reviews in Fish Biology and Fisheries 15: 37-52.

Davidson W, Huang T-K, Fujiki K, Von Schalburg K, Koop B (2009) The sex determining loci and sex chromosomes in the family salmonidae. Sexual Development 3: 78-87.

D'Cotta H, Fostier A, Guiguen Y, Govoroun M, Baroiller J-F (2001) *Aromatase* plays a key role during normal and temperature-induced sex differentiation of tilapia *Oreochromis niloticus*. Molecular Reproduction and Development 59: 265-276.

Deng X, Hiatt JB, Ercan S, Sturgill D, Hillier LW, et al. (2011) Evidence for compensatory upregulation of expressed X-linked genes in mammals, *Caenorhabditis elegans* and *Drosophila melanogaster*. Nature Genetics 43: 1179-1185.

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

Ellegren H (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. Nature Reviews Genetics 12: 157-166.

Eshel O, Shirak A, Weller JI, Hulata G, Ron M (2012) Linkage and physical mapping of sex region on LG23 of Nile tilapia (*Oreochromis niloticus*). G3: Genes|Genomes|Genetics 2: 35-42.

Eshel O, Shirak A, Weller JI, Slossman T, Hulata G, et al. (2011) Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). Animal Genetics 42: 222-224.

Ezaz T, Stiglec R, Veyrunes F, Graves JAM (2006) Relationships between vertebrate ZW and XY sex chromosome systems. Current Biology 16: R736-R743.

Ferguson MW, Joanen T (1982) Temperature of egg incubation determines sex in *Alligator mississippiensis*. Nature 296, 850 – 853.

Foster JW, Graves JAM (1994) An *SRY*-related sequence on the marsupial X chromosome: Implications for the evolution of the mammalian testis-determining gene. Proceedings of the National Academy of Sciences 91: 1927-1931.

Fujimura K, Kocher TD (2011) Tol2-mediated transgenesis in tilapia (*Oreochromis niloticus)*. Aquaculture 319: 342-346.

Giraldez AJ, Cinalli RM, Glasner ME, Enright AJ, Thomson JM, et al. (2005) MicroRNAs regulate brain morphogenesis in zebrafish. Science 308: 833-838.

Golan M, Levavi-Sivan B (2013) Social dominance in tilapia is associated with gonadotroph hyperplasia. General and Comparative Endocrinology 192: 126-135.

Gousset B (1990) European eel (*Anguilla Anguilla* L.) farming technologies in Europe and in Japan: Application of a comparative analysis. Aquaculture 87: 209-235.

Graves JAM (1998) Interactions between *SRY* and *SOX* genes in mammalian sex determination. Bioessays 20: 264-269.

Graves, JAM., Wakefield, MJ., Toder, R. (1998). The origin and evolution of the pseudoautosomal regions of human sex chromosomes. Human molecular genetics, 7(13), 1991-1996.

Graves JAM, Shetty S (2001) Sex from W to Z: evolution of vertebrate sex chromosomes and sex determining genes. Journal of Experimental Zoology 290: 449-462.

Graves, JAM. (2006). Sex chromosome specialization and degeneration in mammals. Cell, 124(5), 901-914.

Griffin DK, Harvey SC, Campos-Ramos R, Ayling LJ, Bromage NR, et al. (2002) Early origins of the X and Y chromosomes: Lessons from tilapia. Cytogenetic and Genome Research 99: 157-163.

Grynnerup AGA, Lindhard A, SØRensen S (2012) The role of *anti-Müllerian hormone* in female fertility and infertility – an overview. Acta Obstetricia et Gynecologica Scandinavica 91: 1252-1260.

Guo H, Ingolia NT, Weissman JS, Bartel DP (2010) Mammalian microRNAs predominantly act to decrease target mRNA levels. Nature 466: 835-840.

Hammerman I, Avtalion R (1979) Sex determination in *Sarotherodon* (tilapia). Theoretical and Applied Genetics 55: 177-187.

Hattori RS, Murai Y, Oura M, Masuda S, Majhi SK, et al. (2012) A Y-linked *anti-Müllerian hormone* duplication takes over a critical role in sex determination. Proceedings of the National Academy of Sciences 109: 2955-2959.

Hattori RS, Strüssmann CA, Fernandino JI, Somoza GM (2013) Genotypic sex determination in teleosts: Insights from the testis-determining *amhy* gene. General and Comparative Endocrinology 192: 55-59.

Hayashi K, Chuva de Sousa Lopes SM, Kaneda M, Tang F, Hajkova P, et al. (2008) MicroRNA biogenesis is required for mouse primordial germ cell development and spermatogenesis. PLoS ONE 3: e1738.

Hong C-S, Park B-Y, Saint-Jeannet J-P (2007) The function of *Dmrt* genes in vertebrate development: It is not just about sex. Developmental Biology 310: 1-9.

Huang CW, Li YH, Hu SY, Chi JR, Lin GH, et al. (2012) Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (*Oreochromis niloticus*). Journal of Animal Science 90: 4266-4279.

Huertas M, Cerdà J (2006) Stocking density at early developmental stages affects growth and sex ratio in the European eel (*Anguilla anguilla*). The Biological Bulletin 211: 286-296.

Ijiri S, Kaneko H, Kobayashi T, Wang D-S, Sakai F, et al. (2008) Sexual dimorphic expression of genes in gonads during early differentiation of a teleost fish, the Nile tilapia *Oreochromis niloticus*. Biology of Reproduction 78: 333-341.

Jamin SP, Arango NA, Mishina Y, Hanks MC, Behringer RR (2003) Genetic studies of the *Amh*/*Mis* signaling pathway for Müllerian duct regression. Mol Cell Endocrinol 211: 15–19.

Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, et al. (2012) A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (Fugu). PLoS Genet 8: e1002798.

Kasahara M, Naruse K, Sasaki S, Nakatani Y, Qu W, et al. (2007) The medaka draft genome and insights into vertebrate genome evolution. Nature 447: 714-719.

Kitano J, Ross JA, Mori S, Kume M, Jones FC, et al. (2009) A role for a neo-sex chromosome in stickleback speciation. Nature 461: 1079-1083.

Kwon, JY., McAndrew, BJ., Penman, DJ. (2001). Cloning of brain aromatase gene and expression of brain and ovarian aromatase genes during sexual differentiation in genetic male and female Nile tilapia *Oreochromis niloticus*. Molecular reproduction and development, 59(4), 359-370.

Kocher TD (2004) Adaptive evolution and explosive speciation: the cichlid fish model. Nature Reviews Genetics 5: 288-298.

Kocher TD, Lee W-J, Sobolewska H, Penman D, McAndrew B (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics 148: 1225-1232.

Kondo M, Hornung U, Nanda I, Imai S, Sasaki T, et al. (2006) Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of medaka. Genome Research 16: 815-826.

Kondo M, Nanda I, Hornung U, Asakawa S, Shimizu N, et al. (2003) Absence of the candidate male sex-determining gene *dmrt1b* (Y) of medaka from other fish species. Current Biology 13: 416-420.

Koop BF, von Schalburg KR, Leong J, Walker N, Lieph R, et al. (2008) A salmonid EST genomic study: genes, duplications, phylogeny and microarrays. BMC Genomics 9: 545.

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

Korol A, Shirak A, Cnaani A, Hallerman EM (2007) Detection and analysis of quantitative trait loci (QTL) for economic traits in aquatic species. Aquaculture Genome Technologies: 169-197.

Koshimizu E, Strüssmann CA, Okamoto N, Fukuda H, Sakamoto T (2010) Construction of a genetic map and development of DNA markers linked to the sex-determining locus in the Patagonian pejerrey (*Odontesthes hatcheri*). Marine Biotechnology 12: 8-13.

Kubiczkova L, Sedlarikova L, Hajek R, Sevcikova S (2012) TGF-beta - an excellent servant but a bad master. Journal of Translational Medicine 10: 183.

Lahav E (1993) Use of sex-reversed females to produce all-male tilapia (*Oreochromis aureus*) fry. Israeli Journal of Aquaculture-Bamidgeh 45: 131-136.

Lee BY, Hulata G, Kocher TD (2004) Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). Heredity 92: 543-549.

Lee BY, Lee WJ, Streelman JT, Carleton KL, Howe AE, et al. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis spp*.). Genetics 170: 237-244.

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

Lee B-Y, Howe A, Conte M, D'Cotta H, Pepey E, et al. (2010) An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. BMC Genomics 11: 278.

Lee B-Y, Kocher TD (2007) Comparative genomics and positional cloning. Aquaculture Genome Technologies: 325-337.

Lester L, Lawson K, Abella T, Palada M (1989) Estimated heritability of sex ratio and sexual dimorphism in tilapia. Aquaculture Research 20: 369-380.

Lewis BP, Shih Ih, Jones-Rhoades MW, Bartel DP, Burge CB (2003) Prediction of mammalian microRNA targets. Cell 115: 787-798.

Li S-C, Tang P, Lin W-C (2007) Intronic microRNA: discovery and biological implications. DNA and Cell Biology 26: 195-207.

Liew WC, Bartfai R, Lim Z, Sreenivasan R, Siegfried KR, et al. (2012) Polygenic sex determination system in zebrafish. PLoS One 7: e34397.

Mair GC, Abucay JS, Skibinski DOF, Abella TA, Beardmore JA (1997) Genetic manipulation of sex ratio for the large-scale production of all-male tilapia, *Oreochromis niloticus*. Canadian Journal of Fisheries and Aquatic Sciences 54: 396-404.

Mair GC, Scott AG, Penman DJ, Beardmore JA, Skibinski DOF (1991) Sex determination in the genus *Oreochromis* .1. Sex reversal, gynogenesis and triploidy in *Oreochromis niloticus* (L). Theoretical and Applied Genetics 82: 144-152.

Majumdar KC, McAndrew BJ (1986) Relative DNA content of somatic nuclei and chromosomal studies in three genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe *Tilapiini* (Pisces, Cichlidae). Genetica 68: 175-188.

Mank JE, Promislow DEL, Avise JC (2006) Evolution of alternative sex-determining mechanisms in teleost fishes. Biological Journal of the Linnean Society 87: 83-93.

Manolio TA, Collins FS, Cox NJ, Goldstein DB, Hindorff LA, et al. (2009) Finding the missing heritability of complex diseases. Nature 461: 747-753.

Martínez P, Bouza C, Hermida M, Fernández J, Toro MA, et al. (2009) Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). Genetics 183: 1443-1452.

Matsuda M (2005) Sex determination in the teleost medaka, *Oryzias latipes*. Annual Reviews Genetics 39: 293-307.

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

Matsuda M, Shinomiya A, Kinoshita M, Suzuki A, Kobayashi T, et al. (2007) *DMY* gene induces male development in genetically female (XX) medaka fish. Proceedings of the National Academy of Sciences 104: 3865-3870.

McFarlane L, Wilhelm D (2010) Non-Coding RNAs in mammalian sexual development. Sexual Development 3: 302-316.

McKaye KR, Stauffer J, Van Der Berghe E, Vivas R, Lopez Perez L, et al. (2002) Behavioral, morphological and genetic evidence of divergence of the Midas cichlid species complex in two Nicaraguan crater lakes. Cuadernos de Investigación de la UCA 12: 19-47.

Mishima T, Takizawa T, Luo S-S, Ishibashi O, Kawahigashi Y, et al. (2008) MicroRNA (miRNA) cloning analysis reveals sex differences in miRNA expression profiles between adult mouse testis and ovary. Reproduction 136: 811-822.

Mishina Y, Whitworth DJ, Racine C, Behringer RR. 1999. High specificity of Mullerian-inhibiting substance signaling in vivo. Endocrinology 140: 2084–2088

Mittwoch U (2006) Sex is a threshold dichotomy mimicking a single gene effect. Trends in Genetics 22: 96-100.

Morgan C, Bale T (2012) Sex differences in microRNA regulation of gene expression: no smoke, just miRs. Biology of Sex Differences 3: 22.

Morinaga, C., Saito, D., Nakamura, S., Sasaki, T., Asakawa, S., et al (2007). The hotei mutation of medaka in the *anti-Müllerian hormone receptor* causes the dysregulation of germ cell and sexual development. Proceedings of the National Academy of Sciences, 104(23), 9691-9696.

Muller, HJ. (1914). A gene for the fourth chromosome of Drosophila. Journal of Experimental Zoology, 17(3), 325-336.

Myosho T, Otake H, Masuyama H, Matsuda M, Kuroki Y, et al. (2012) Tracing the emergence of a novel sex-determining gene in medaka, *Oryzias luzonensis*. Genetics 191: 163-170.

Nagai T, Takehana Y, Hamaguchi S, Sakaizumi M (2008) Identification of the sex-determining locus in the Thai medaka, *Oryzias minutillus*. Cytogenetic and Genome Research 121: 137-142.

Nanda I, Kondo M, Hornung U, Asakawa S, Winkler C, et al. (2002) 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.

Nelson JS (1994) Fishes of the World. Wiley, New York, NY: 600.

Nelson JS (2006) Fishes of the World: Wiley. com.

Opferman J (2007) Apoptosis in the development of the immune system. Cell Death & Differentiation 15: 234-242.

Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, et al. (2013) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD Sequencing. PLoS ONE 8: e68389.

Palti Y, Shirak A, Cnaani A, Hulata G, Avtalion RR, et al. (2002) Detection of genes with deleterious alleles in an inbred line of tilapia (*Oreochromis aureus*). Aquaculture 206: 151-164.

Papagiannakopoulos T, Shapiro A, Kosik KS (2008) MicroRNA-21 targets a network of key tumor-suppressive pathways in glioblastoma cells. Cancer Research 68: 8164-8172.

Plasterk RHA (2006) Micro RNAs in animal development. Cell 124: 877-881.

Pomiankowski A, Nöthiger R, Wilkins A (2004) The evolution of the *Drosophila* sex-determination pathway. Genetics 166: 1761-1773.

Poonlaphdecha S, Pepey E, Canonne M, de Verdal H, Baroiller J-F, et al. (2013) Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *dmrt1* and *amh* expressions. General and Comparative Endocrinology 193: 234-242.

Pruginin Y, Rothbard S, Wohlfarth G, Halevy A, Moav R, et al. (1975) All-male broods of Tilapia nilotica × T. aurea hybrids. Aquaculture 6: 11-21.

Raymond CS, Kettlewell JR, Hirsch B, Bardwell VJ, Zarkower D (1999) Expression of *Dmrt1* in the genital ridge of mouse and chicken embryos suggests a role in vertebrate sexual development. Developmental Biology 215: 208-220.

Rens W, Grützner F, O'Brien PC, Fairclough H, Graves JAM JA, et al. (2004) Resolution and evolution of the duck-billed platypus karyotype with an X1Y1X2Y2X3Y3X4Y4X5Y5 male sex chromosome constitution. Proceedings of the National Academy of Sciences of the United States of America 101: 16257-16261.

Rice WR (1987) Genetic hitchhiking and the evolution of reduced genetic activity of the Y sex chromosome. Genetics 116: 161-167.

Rodríguez-Marí A, Cañestro C, BreMiller RA, Nguyen-Johnson A, Asakawa K, et al. (2010) Sex reversal in zebrafish *fancl* mutants is caused by Tp53-mediated germ cell apoptosis. PLoS Genet 6: e1001034.

Rosenstein S, Hulata G (1994) Sex reversal in the genus *Oreochromis*: optimization of feminization protocol. Aquaculture Research 25: 329-339.

Ross JA, Urton JR, Boland J, Shapiro MD, Peichel CL (2009) Turnover of sex chromosomes in the stickleback fishes (*Gasterosteidae*). PLoS genetics 5: e1000391.

Rougeot C, Prignon C, Kengne CVN, Melard C (2008) Effect of high temperature during embryogenesis on the sex differentiation process in the Nile tilapia, *Oreochromis niloticus*. Aquaculture 276: 205-208.

Saito D, Tanaka M (2009) Comparative aspects of gonadal sex differentiation in medaka: a conserved role of developing oocytes in sexual canalization. Sexual Development 3: 99-107.

Schaible UE, Winau F, Sieling PA, Fischer K, Collins HL, et al. (2003) Apoptosis facilitates antigen presentation to T lymphocytes through MHC-I and CD1 in tuberculosis. Nature Medicine 9: 1039-1046.

Schartl M (2004) Sex chromosome evolution in non-mammalian vertebrates. Current Opinion in Genetics & Development 14: 634-641.

Schliewen UK, Tautz D, Pääbo S (1994) Sympatric speciation suggested by monophyly of crater lake cichlids. Nature 368: 629-632.

Schutt C, Nothiger R (2000) Structure, function and evolution of sex-determining systems in Dipteran insects. Development 127: 667-677.

Shirak A, Palti Y, Cnaani A, Korol A, Hulata G, et al. (2002) Association between loci with deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). Journal of Heredity 93: 270-276.

Shirak A, Seroussi E, Cnaani A, Howe AE, Domokhovsky R, et al. (2006) *Amh* and *Dmrta2* genes map to tilapia (*Oreochromis spp*.) linkage group 23 within quantitative trait locus regions for sex determination. Genetics 174: 1573-1581.

Siegfried KR, Nüsslein-Volhard C (2008) Germ line control of female sex determination in zebrafish. Developmental Biology 324: 277-287.

Sinclair AH, Berta P, Palmer MS, Hawkins JR, Griffiths BL, et al. (1990) A Gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature 346: 240-244.

Smith CA, Katz M, Sinclair AH (2003) *DMRT1* is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. Biology of Reproduction 68: 560-570.

Smith CA, Roeszler KN, Ohnesorg T, Cummins DM, Farlie PG, et al. (2009) The avian Z-linked gene *DMRT1* is required for male sex determination in the chicken. Nature 461: 267-271.

Smrzka OW, Delgehyr N, Bornens M (2000) Tissue-specific expression and subcellular localisation of mammalian *δ-tubulin*. Current Biology 10: 413-416.

Takehana Y, Hamaguchi S, Sakaizumi M (2008) Different origins of ZZ/ZW sex chromosomes in closely related medaka fishes, *Oryzias javanicus* and *O. hubbsi.* Chromosome Research 16: 801-811.

Takehana Y, Naruse K, Hamaguchi S, Sakaizumi M (2007) Evolution of ZZ/ZW and XX/XY sex-determination systems in the closely related medaka species, *Oryzias hubbsi* and *O. dancena*. Chromosoma 116: 463-470.

Thomson DW, Bracken CP, Goodall GJ (2011) Experimental strategies for microRNA target identification. Nucleic Acids Research 39: 6845-6853.

Tong S-K, Hsu H-J, Chung B-c (2010) Zebrafish monosex population reveals female dominance in sex determination and earliest events of gonad differentiation. Developmental Biology 344: 849-856.

Tuan PA, Mair GC, Little DC, Beardmore JA (1999) Sex determination and the feasibility of genetically male tilapia production in the Thai-Chitralada strain of *Oreochromis niloticus* (L.). Aquaculture 173: 257-269.

Uchida D, Yamashita M, Kitano T, Iguchi T (2002) Oocyte apoptosis during the transition from ovary-like tissue to testes during sex differentiation of juvenile zebrafish. Journal of Experimental Biology 205: 711-718.

Van Doorn G, Kirkpatrick M (2007) Turnover of sex chromosomes induced by sexual conflict. Nature 449: 909-912.

Ventura T, Manor R, Aflalo ED, Weil S, Raviv S, et al. (2009) Temporal silencing of an androgenic gland-specific insulin-like gene affecting phenotypical gender differences and spermatogenesis. Endocrinology 150: 1278-1286.

Volff JN, Nanda I, Schmid M, Schartl M (2007) Governing sex determination in fish: Regulatory putsches and ephemeral dictators. Sexual Development 1: 85-99.

Volff J-N, Schartl M (2001) Variability of genetic sex determination in poeciliid fishes. Genetica 111: 101-110.

Wallace B, Wallace H (2003) Synaptonemal complex karyotype of zebrafish. Heredity 90: 136-140.

Weller JI, Ron M (2011) Invited review: Quantitative trait nucleotide determination in the era of genomic selection. Journal of Dairy Science 94: 1082-1090.

Wienholds E, Kloosterman WP, Miska E, Alvarez-Saavedra E, Berezikov E, et al. (2005) MicroRNA expression in zebrafish embryonic development. Mechanisms of Development 122: S149-S150.

Wohlfarth GW (1994) The unexploited potential of tilapia hybrids in aquaculture. Aquaculture and Fisheries Management 25: 781-788.

Wohlfarth GW, Hulata GI (1981) Applied Genetics of Tilapias. ICLARM Studies and Reviews, 6**:**1-26.

Yan B, Guo J-T, Zhao L-H, Zhao J-L (2012) microRNA expression signature in skeletal muscle of Nile tilapia. Aquaculture 364–365: 240-246.

Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, et al. (2012) An immune-related gene evolved into the master sex-determining gene in Rainbow Trout, *Oncorhynchus mykiss*. Current Biology 22: 1423-1428.

Yoshimoto S, Okada E, Umemoto H, Tamura K, Uno Y, et al. (2008) A W-linked DM-domain gene, *DM-W*, participates in primary ovary development in *Xenopus laevis*. Proceedings of the National Academy of Sciences 105: 2469-2474.

תקציר

אמנונים הם דגים בעלי חשיבות כלכלית רבה בחקלאות הימית. ממשק גידול האמנונים מתבסס על גידול אוכלוסיות חד-זוויגיות של זכרים. ישנן מספר שיטות המיושמות כיום לקבלת אוכלוסיות אילו, אך קיים קושי לתחזק אותן בקנה מידה מסחרי לאורך זמן. לדוגמא, היפוך זוויג ע"י טיפול הורמונלי הינו בעייתי לשימוש הן למגדלים והן לסביבה, הכלאות בין פרטים מסוימים מניבים הטלות של זכרים (~100%), אך מאגר ההורים מזדהם לאורך זמן. לכן, להבנת המנגנון לקביעת זוויג באמנונים חשיבות מדעית ומסחרית רבה.

ההשערה שמערכת כרומוזומי הזוויג של אמנון יאור היא XX/XY מבוססת בעיקר על תוצאות הכלאות עם פרטים שעברו היפוך זוויג הורמונלי. לדוגמא, מהכלאה בין זכרים הפוכי זוויג (ΔXX) עם נקבות רגילות (XX) מתקבלות נקבות (~ 100%). עבודות קודמות הראו כי קביעת זוויג באמנון יאור מבוקרת על ידי גורמים גנטיים שונים ופעולת גומלין ביניהם לגורמים סביבתיים, ולכן הוצע לחקור את קביעת הזוויג באמנונים כתכונה כמותית. אתרים גנומיים המשפיעים על תכונות כמותיות נקראים QTL (Quantitative Trait Loci). עבודות קודמות באמנונים מיפו QTL לקביעת זוויג בקבוצות תאחיזה 1 ו-3 על משפחות וזני אמנונים שונים. מספר סמנים הראו קשר לשינוי ביחס זוויגים בקבוצות תאחיזה 2, 6 ו-23 באמנון ירדן. בגלל רווח אמינות גדול של ה-QTL האזורים הגנומיים המדווחים לקביעת זוויג היו נרחבים וכללו עשרות גנים ויותר. בעבודה קודמת מצאו גנים עם רמות ביטוי שונות באופן מובהק בין זכרים לנקבות בגונדות לא ממוינות בימים 9-10 לאחר הפריה. בנוסף, מיקרו-רנ"א הם גורמי בקרה שמורים בקרב אאוקריוטים בעלי תפקיד חשוב בתהליכים ביולוגיים רבים במהלך ההתפתחות, כגון גדילת תאים, התמיינות ואפופטוזיס. יתר על כן, מחקרים במינים שונים מצאו רמות ביטוי שונות של מיקרו-רנ"א בין הזוויגים בגונדות ובעובר.

מטרות מחקר זה היו מיפוי עדין שלQTL לקביעת זוויג אשר דווחו בספרות וזיהוי גנים ומיקרו-רנ"א המשפיעים על קביעת זוויג בשלבים מוקדמים של התפתחות עוברית של אמנון היאור.למיפוי האזורים לקביעת זוויג, קבוצת אחים מלאים של אמנון היאור חולקה לשלוש קבוצות שקיבלו טיפולים שונים: (1) ללא טיפול (2) היפוך זוויג לנקבות ע"י di-ethylstilbestrol ו-(3) היפוך זוויג לזכרים ע"י 17a-methyl-testosterone. הקבוצה הלא מטופלת נסרקה באמצעות סמנים גנטיים המייצגים QTLs לקביעת זוויג בקבוצות תאחיזה 1, 2, 3, 6 ו- 23. הסמן שנמצא עם הקשר החזק ביותר לזוויג היה *UNH898* הממופה לקבוצת תאחיזה 23 (χ^2, $p=8.6x10^{-5}$). אלל 276 נמצא כמעט בלעדית בזכרים, ולכן השערת העבודה הייתה כי אלל זה ייחודי לזכרים. כדי לבדוק השערה זו נקבע גנוטיפ לסמן *UNH898* בפרטים הפוכי זוויג בקבוצת טיפול 2 ו- 3 ונעשו הכלאות בין פרטים מסוימים שנבחרו עפ"י הגנוטיפ שלהם. בהכלאות בין פרטים ללא אלל 276 התקבלו % 97.5 נקבות, והכלאה בין פרטים ששניהם הטרוזיגוטיים לאלל 276 התקבלו 81 זכרים ו-30 נקבות (יחס של 1:3 כצפוי ממערכת קביעת זוויג XX/XY). כך שהקשר בין הזוויג לגנוטיפ בסמן *UNH898* נמצא מובהק (χ^2, p= $2.5x10^{-18}$). בנוסף, קביעת גנוטיפ של שמונה סמנים גנטיים החובקים את *UNH898* אפשרה מיפוי עדין של QTL על קבוצת תאחיזה 23 לרווח אמינות של 16-21 סנטי-מורגן.

מיפוי פיזי ל-QTL על קבוצת תאחיזה 23 נעשה על-ידי קביעת גנוטיפ של 90 פרטים ממשפחה מתפצלת לסמנים גנטיים נוספים. נמצא שהאזור לקביעת-זוויג נמצא בין הסמנים *GM597* ל-*ARO124* ב-"פיגום" הגנומי 101 בין 990,577 ל-2,468,000 בסיסים. שנים-עשר הסמנים הגנטיים הסמוכים באזור זה היו הומוזיגוטים בנקבות, וגם

הומוזיגוטים לאלל השני או הטרוזיגוטיים אצל זכרים. סמנים גנטיים החובקים אזור זה היו הטרוזיגוטיים בשתי נקבות וכך אפשרו למפות את ה-QTL על קבוצת תאחיזה 23 למקטע פיזי של 1.47אלפי בסיסים. אזור גנומי זה מכיל 51 גנים מועמדים לקביעת זוויג כולל הגן anti-Müllerian hormone (*amh*) בעל פעילות ידועה בתהליכי קביעת זוויג. היכולת לצמצם אזור זה למקטע קטן יותר דורשת דיגום של אלפי דגים כדי לזהות רקומביננטים בין הגן לזוויג וה-QTL.

ריצוף גנום האמנון אפשר חיפוש גנים ומיקרו-רנ"א על פני כל הגנום המשפיעים על קביעת זוויג. מטרת העבודה היתה לזהות גנים ומיקרו-רנ"א בעלי רמות ביטוי שונות בין זכרים לנקבות בימים 2, 5 ו -9 לאחר ההפריה המייצגים שלבים מוקדמים בהתפתחות עוברית. העוברים נלקחו ממשפחות חד-זוויגיות שהתקבלו ע"י הפריה מלאכותית של נקבות אמנון יאור עם זכרים הפוכי זוויג (ΔXX) או זכרים 'YY' (פרטים המתקבלים אחרי מס' דורות הכלאה עם פרטים הפוכי זוויג לקבלת פרטים המייצרים זכרים בלבד). 56 דגימות ביולוגיות עברו קשירה עם שבבי-דנ"א, ושימשו בסיס לריצוף מקטעי רנ"א קצרים (עד 45 בסיסים). השבב ייצג את מכלול הגנים של אמנון היאור מאחר שתוכנן על בסיס ניבוי גנים מגנום האמנון וספריות ביטוי מרקמות שונות. השבב הורכב מ-43,803 רצפים באורך של 60 בסיסים, ומניתוח תוצאות הקשירה וההבדלים בין הדוגמאות נמצאו 59 גנים עם רמות ביטוי השונות באופן מובהק בין זכרים לנקבות בשיעור גילוי שקרי של 5 %. בנוסף נמצאו 831 מיקרו-רנ"א מתוכם תשעה עם דפוסי ביטוי שונה (מעל 4 סטיות תקן) בין הזוויגים ובשיעור גילוי שקרי של 1.7 %.

הגנים עם רמות הביטוי גבוהות בנקבות היו: *Carbonyl Reductase-Like 20β-Hydroxysteroid Dehydrogenase* (*cr/20β-hsd*; $p=1\times10^{-17}$) ו- *Reticulon-4-interacting protein 1 homolog* ($p=1\times10^{-17}$), *Zinc-finger BED Domain- containing 3* ($p=1\times10^{-12}$) *Tetraspanin-8* ($p=1\times10^{-16}$) ואילו הגנים עם רמות הביטוי הגבוהות בזכרים היו: *Inositol monophosphatase 1* ($p=1\times10^{-14}$). רמות הביטוי של *amh* ו-*amh* ($p=1\times10^{-9}$). נבדקו ברקמת מוח, כבד וגונדות בזכרים ונקבות ביום 75 לאחר הפריה, ונמצא כי *cr/20β-hsd* מתבטא ביתר בשחלה ואילו ל- *amh* רמת ביטוי גבוהה במוח ובאשך באופן מובהק בזכרים.

שמונה גנים נדגמו לבדיקת אימות הממצאים בשיטת PCR כמותי ונמצא מתאם של 0.8 ביחס הביטוי בין ניסוי הקשירה לשבב דנ"א ו-PCR כמותי. עם זאת, ניתוח עקומת-התכה של ה- PCR הכמותי הצביע על ריבוי עותקים של גנים מסוימים באחד הזוויגים. עבודה נוספת של PCR כמותי על בסיס דנ"א של זכרים ונקבות הראתה שלארבעה מתוך שמונה גנים *amh*, *cr/20β-hsd*, *tspan8* ו-*gpa33* יש הבדל מובהק במספר העותקים בין המינים. בכל ארבעת הגנים היה קשר חיובי בין מספר העותקים ורמת ביטוי בזוויג מסוים.

amh היה גן מועמד עיקרי לקביעת זוויג במספר רמות: מיקומו ב-QTL לקביעת זוויג בקבוצת תאחיזה 23, פעילותו במסלולים לקביעת זוויג באורגניזמים אחרים, וקיום מספר עותקים גבוה יותר בגנום הזכרי ורמות ביטוי גבוהות בעובר, מוח ואשך בזכרים. בעבודת ריצוף פרטנית לגן זה זוהה עותק מוכפל של הגן שהינו ייחודי בזכרים. העותק הזה נקרא *amhy*, וההבדל העיקרי מהרצף המשותף לשני הזוויגים הוא שאקסון 7, המכיל אתר קישור לרצפטור *transforming growth factor beta*, הינו קצר יותר ב- 233 בסיסים ומכאן שחסר את היכולת לקודד חלבון מלא. בנוסף, נמצא כי *amhy* מתבטא באשך, אך לא בשחלה.

שישה מתוך תשעה מיקרו-רנ"א התבטאו ביתר בזכרים בשלבי התפתחות מוקדמים בעובר. שני מיקרו-רנ"א נמצאו באופן עצמאי בשתיים משלוש נקודות הזמן ואילו שניים אחרים נמצאו בכל שלוש נקודות הזמן. קצה '3 של 59 הגנים אשר נמצאו עם רמות ביטוי שונות בין זכרים לנקבות בניסוי הקשירה לשבב הדנ"א, נבדקו כגני מטרה לתשעה מיקרו-רנ"א אשר

באו לידי ביטוי באופן מבדיל בין המינים. רק pma-mir-4585 המתבטא ביתר בזכרים הראה קשר הפוך באופן מובהק לששת גני המטרה הצפויים שלו שהציגו ביטוי חסר בזכרים. שניים מתשעת המיקרו-רנ"א (mir-21 ו-mir-218) המתבטאים ביתר בזכרים, הינם שמורים בקרב בעלי חוליות וכך ניתן היה לקבוע את זהות הגנים המארחים שלהם. התקבלו שני גנים שעל פי הדיווח בספרות הם פעילים במסלולים הקשורים למערכת הזוויג הזכרית בביטוי באשך וייצור תאי זרע. מידע זה הינו בהתאמה לכך שמיקרו-רנ"א והגנים המארחים מבוטאים יחד.

לסיכום, מחקר זה מדווח על מיפוי עדין של QTL לקביעת זוויג על קבוצת תאחיזה 23, גילוי דפוסי ביטוי מובדלים של גנים ומיקרו-רנ"א בין הזוויגים, והעשרה לגנים בעלי מספר עותקים בגנום. זו העבודה הראשונה בה נמצאה הכפלת הגן *amh* בעותק ייחודי שנמצא בזכרים ושהמסלולים לקביעת-זוויג מתחילים זמן קצר לאחר ההפריה. בנוסף, נמצאה קבוצה גדולה של גנים בעלי ביטוי שונה בין הזכרים ונקבות המעורבים באפופטוזיס ומהווים גני מטרה למיקרו-רנ"א המתבטאים באופן מובהק בזוויג מסוים. שילוב הממצאים שנאספו בעבודה בכלים שונים מצביע כי הגנים *amh*, ו-*cr/20β-hsd* ותהליך האפופטוזיס מעורבים בקביעת זוויג באמנונים בדומה למודל קביעת זוויג שהוצע לדג-זברה.

תודות

עבודה זו נעשתה בעזרת אנשים רבים לכל אורך הדרך ועל כך נתונה להם תודתי.

ברצוני להודות למנחים, פרופ' מיכה רון ופרופ' גידי חולתא אשר היוו עבורי דוגמא אישית ומקצועית והעניקו לי את ההזדמנות והחופש לעשות את עבודת הדוקטוראט בכיוון בו בחרתי.

תודה לדר' אנדריי שיראק על הלימוד והייעוץ לאורך העבודה.

תודה לדר' איל סרוסי ודר' יהודה ולר על העזרה בחלקים שונים של העבודה.

תודה לדר' גיורא גליק על ההקשבה והחברות.

תודה לליאור דור על החיוך הקבוע ועל שתמיד מוכן לעזור.

ותודה מיוחדת לבעלי היקר, עדי, אילולא עידודו ותמיכתו לא הייתה מתחילה ולבטח לא מסתיימת עבודה זו.

עבודה זו נעשתה בהדרכתם של פרופ׳ מיכה רון ופרופ׳ גדעון חולתא.

זיהוי אתרים בגנום, גנים ומיקרו-רנ"א המעורבים בקביעת זוויג והתמיינות באמנון יאור

חיבור לשם קבלת תואר דוקטור לפילוסופיה

מאת אורלי אשל

הוגש לסנט האוניברסיטה העברית, בירושלים

דצמבר, 2013