Genomic analysis detects polygenic sex determination and signs of early gonad differentiation in zebrafish

Liew, Woei Chang

2014

Liew, W. C. (2014). Genomic analysis detects polygenic sex determination and signs of early gonad differentiation in zebrafish. Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/58908>

<https://doi.org/10.32657/10356/58908>

**GENOMIC ANALYSIS DETECTS POLYGENIC SEX
DETERMINATION AND SIGNS OF EARLY GONAD
DIFFERENTIATION IN ZEBRAFISH**

LIEW WOEI CHANG

School of Biological Sciences

2014

**GENOMIC ANALYSIS DETECTS POLYGENIC SEX
DETERMINATION AND SIGNS OF EARLY GONAD
DIFFERENTIATION IN ZEBRAFISH**

LIEW WOEI CHANG

School of Biological Sciences

A thesis submitted to the Nanyang Technological University
in partial fulfillment of the requirement for the degree of
Doctor of Philosophy

2014

Acknowledgements

I am grateful to my supervisor, Prof László Orbán, for his guidance and support during my PhD journey that has molded me into a better scientist and man. His warmness and approachable character makes him my source for invaluable advice both at the lab bench and outside the lab.

I am also thankful to my colleagues (both present and past) who have come alongside me on this journey. The insightful exchanges and collaborations provided me with new perspectives on many areas of science. Their friendships are also a treasure that I will keep forever.

Most importantly, I would like to express my heartfelt gratitude to my family members, especially my parents and fiancée. Their continuous support, understanding and love are one of the main sources of motivation for me.

I would also like to acknowledge the generous funding from Temasek Life Sciences Laboratory, Agri-Food & Veterinary Authority of Singapore and National Research Foundation.

Last but not least, I like to thank God for His faithfulness and blessing to me in my life and research for all these year.

Table of Contents

Acknowledgements.....	i
Table of Contents.....	ii
List of Figures	v
List of Tables	vii
List of Appendixes	viii
List of Publications and Presentations.....	ix
Abbreviations.....	xii
Abstract.....	xiv
1. Introduction.....	1
1.1 Zebrafish: A teleost model organism	1
1.2 Sex determination.....	2
1.3 Genetic (genotypic) sex determination.....	3
1.4 Contradicting data on the zebrafish SD mechanism.....	4
1.4.1 Cytogenetic methods to search for sex chromosomes.....	4
1.4.2 Detection of the CSD system by whole genome manipulation	7
1.4.3 Breeding sex-reversed fish to resolve the SD mechanism.....	11
1.5 Environmental factors exert secondary effects on zebrafish SD.....	12
1.6 Zebrafish gonads go through a mandatory “juvenile ovary” phase	15
1.7 Multiple genetic pathways are involved in gonad differentiation.....	16
1.8 Objectives of this study.....	19

2	Materials and Methods.....	21
2.1	Fish stocks and tail fin samples	21
2.2	Fish husbandry	21
2.3	Visual sexing of zebrafish.....	22
2.4	Selective breeding experiment	23
2.5	FluoMEP assay	23
2.6	Array comparative genomic hybridization.....	24
2.7	Validation of CNVRs by real-time quantitative PCR.....	26
2.8	RNA extraction.....	27
2.9	Expression microarray	28
2.10	Expression microarray data analysis	30
2.11	High throughput real-time quantitative PCR array for expression analysis	32
2.12	qPCR array data analysis.....	33
3	Results	35
3.1	Zebrafish has wide-ranging inter-family sex ratios	35
3.2	Sex-linked markers identified by FluoMEP in guppy and rosy barb, but not in zebrafish.....	36
3.3	The absence of a universal sex-linked CNVR supports lack of differentiated sex chromosomes in zebrafish.....	39
3.4	Highly similar progeny sex ratios from repeated pairwise crossings indicate the presence of a GSD system in zebrafish.....	44
3.5	Response to selective full-sib pairwise breeding was in accordance with PSD system.....	45
3.6	A large number of transcripts showed dimorphic expression level during the gonadal differentiation period	50
3.7	Enrichment of functional gene sets from GO and KEGG databases	57

3.8	Expression profile of a selected set of genes revealed transcriptomic differences in zebrafish gonads from 28 to 42 dpf.....	61
3.9	The expression pattern of a selected gonadal gene set indicated that sex determination in zebrafish started prior to 7 dpf.....	69
4	Discussion	72
4.1	Zebrafish sex is determined by genetic factor(s).....	72
4.2	Zebrafish is not a CSD species.....	73
4.3	The zebrafish has a PSD system.....	79
4.4	A concept for variable sex ratios observed in PSD species.....	80
4.5	Zebrafish gonad differentiation starts before the juvenile ovary phase.....	85
4.6	Summary and future outlook.....	88
5	References	92

List of Figures

Figure 1: Inferring the type of sex chromosomal system by progeny sex ratios as a result of genome manipulation or artificial sex reversal.....	9
Figure 2: Shift in the balance of pro-male and pro-female pathways will determine the direction of gonad differentiation in zebrafish.	19
Figure 3: Schematic diagram of the FluoMEP principle.....	25
Figure 4: Databases used to custom design Zv9 expression microarray.....	30
Figure 5: Sex ratios of 62 zebrafish families showed a wide-ranging profile.	36
Figure 6: Male-specific FluoMEP markers from guppy.	38
Figure 7: Male-specific FluoMEP markers from rosy barb.	39
Figure 8: Genomic map of the zebrafish copy number variation regions.....	41
Figure 9: The relative copy number of three CNVRs chosen for gDNA-qPCR validation.....	43
Figure 10: Repeated crossings of the same mating pair produced similar progeny sex ratios.....	45
Figure 11: Selective full-sib pairwise breeding of H3 line maintained female-biased sex ratio and decreased sex ratio variation.	47
Figure 12: Selective full-sib pairwise breeding of the male-biased D line.	48
Figure 13: Sample collection scheme for expression analysis of gonad differentiation by microarray.	51
Figure 14: PCA plot of the microarray data showed distinct divergence between the gonadal transcriptomes of strong (green) and weak (orange) vasa-EGFP fluorescence groups.	52
Figure 15: Comparison of up-regulated transcripts in strong or weak fluorescence groups between 28 to 105 dpf.....	55
Figure 16: Validation of expression microarray result by qPCR.	56

Figure 17: Comparison of GO terms found to be significantly up-regulated in the strong or weak fluorescence groups from all time points.....	58
Figure 18: Comparison of KEGG pathways up-regulated in strong or weak fluorescence groups.....	60
Figure 19: Expression profiling of zebrafish genes associated with gonad development revealed an intermediate group at 28 and 35 dpf.	64
Figure 20: Expression levels of “pro-male” and “pro-female” genes at 35 dpf confirmed the presence of an intermediate fluorescence group.....	66
Figure 21: Expression profiles of two “pro-male” and “pro-female” genes at 28 dpf.	68
Figure 22: Expression profiling of a selected set of genes during larval stages in zebrafish showed the divergence of gonadal transcript levels starting from 7 dpf.....	70
Figure 23: Family sex ratio profile of CSD and PSD species.....	77
Figure 24: A simplified mechanistic model for PSD system based on the involvement of four autosomal genes.....	82
Figure 25: Influence of zebrafish parental ‘SD allele sets’ on offspring sex indicates multiple loci are involved.....	84

List of Tables

Table 1: Published data from induced parthenogenesis experiments on zebrafish yield sex ratios incompatible with one or both CSD systems.....	10
Table 2: Genomic location of the randomly chosen CNVRs for gDNA-qPCR validation and their HX12 aCGH microarray data for different families.....	42
Table 3: List of the 69 DETs consistently up-regulated in strong (vs. weak) fluorescence groups from 28 to 105 dpf zebrafish.....	54
Table 4: GO terms and KEGG pathways enriched in the list of 69 DETs up-regulated in strong fluorescence groups from 28 to 105 dpf zebrafish.....	55
Table 5: GO terms that were up-regulated in the strong fluorescence group from 28 to 105 dpf.....	59
Table 6: KEGG pathways with consistent expression pattern from 35 to 105 dpf zebrafish.....	60
Table 7: Genes and transcripts selected for qPCR array from the microarray result and published studies.....	62
Table 8: Mean relative quantity of the 20 genes measured by qPCR array for 28, 35 and 42 dpf zebrafish juveniles.....	63
Table 9: Mean relative quantity of the 20 genes measured by qPCR array for 7, 14 and 21 dpf zebrafish larvae.....	71

List of Appendixes

Appendix 1: Primer sequences of sex-linked FluoMEP markers for guppy and rosy barb.	110
Appendix 2: The distribution of aCGH probes on each chromosome	111
Appendix 3: Primer sequences used for CNVR gDNA-qPCR validation and their genomic location.....	112
Appendix 4: Gene details and primer sequences used for expression microarray validation and qPCR array	113
Appendix 5: Family trees and progeny sex ratios of the two zebrafish lines used for selective breeding.....	114

List of Publications and Presentations

Journal publications:

Ravi P, Jiang JH, **Liew WC** & Orban L. (2014) Small-scale transcriptomics reveals differences among gonadal stages in Asian seabass (*Lates calcarifer*). *Reprod Biol Endocrinol* **12**:5.

Liew WC & Orban L. (2014) Zebrafish sex: A complicated affair. *Brief Funct Genomics* **13**(2):172.

Liew WC, Bartfai R, Lim Z, Sreenivasan R, Siegfried KR & Orban L. (2012) Polygenic sex determination system in zebrafish. *PloS One* **7**(4): e34397.

Chang A, **Liew WC**, Chuah A, Lim Z, Lin Q & Orban L. (2007) FluoMEP: A new genotyping method combining the advantages of randomly amplified polymorphic DNA and amplified fragment length polymorphism. *Electrophoresis* **28**(4): 525-534.

Yue GH, **Liew WC** & Orban L. (2006) The complete mitochondrial genome of a basal teleost, the Asian arowana (*Scleropages formosus*, *Osteoglossidae*). *BMC Genomics* **7**:242.

Conference presentations (presenter underlined):

Liew WC, Preethi R & Orban L. (2013) Expression analysis of zebrafish gonad development reveals early divergence *Plant and Animal Genome Asia*. 17-19 Mar 2013, Singapore (Poster).

Liew WC, Jiang J, Sreenivasan R, Ravi P, Thevasagayam NM & Orban L. (2012) Polygenic sex determination and conserved features of gonad differentiation in zebrafish. *6th International Symposium on the Biology of Vertebrate Sex Determination*. 23-27 Apr 2012, USA (Oral).

Liew WC, Bartfai R, Lim Z, Sreenivasan R & Orban L. (2011) Zebrafish sex is genetically determined without differentiated sex chromosomes. *9th International Symposium on Reproductive Physiology of Fish*. 9-14 Aug 2011, India (Poster).

Liew WC, Bartfai R, Lim Z, Sreenivasan R & Orban L. (2010) Zebrafish sex is genetically determined without differentiated sex chromosomes. *Singapore Zebrafish Symposium*, 8 Nov 2010, Singapore (Oral).

Liew WC, Bartfai R, Kwan HY, Thevasagayam NM, Maurer-Stroh S & Orban L. (2009) Zebrafish sex determination appears to be genetic. *5th International Symposium on the Biology of Vertebrate Sex Determination*, 20-24 Apr 2009, USA (Poster).

Liew WC, Chang A, Ye MS, Yap KC & Orban L. (2007) Using mtSNPs to identify the sex of mouthbrooding parent in Asian arowana (*Scleropages formosus*). *44th annual meeting of the Animal Behavior Society*, 21-25 Jul 2007, USA (Poster).

Liew WC, Chang A, Ang WK, Yap KC & Orban L. Genetic relatedness of Asian arowana varieties: Introducing a new strain from Indonesia. *Genetics in Aquaculture IX*, 26-30 Jun 2006, France (Poster).

Abbreviations

aCGH	Array Comparative Genomic Hybridization
AFLP	Amplified Fragment Length Polymorphism
Cq	Quantification cycle
CSD	Chromosomal Sex Determination
CV%	Coefficient of Variation
DET	Differentially Expressed Transcripts
DH	Double Haploid
dpf	Day Post Fertilization
ESD	Environmental Sex Determination
EST	Expressed Sequence Tag
FDR	False Discovery Rate
FluoMEP	Fluorescent Motif Enhanced Polymorphism
gDNA	Genomic Deoxyribonucleic Acid
GO	Gene Ontology
GSD	Genetic/genotypic Sex Determination
HC	Hierarchical Clustering
KEGG	Kyoto Encyclopedia of Genes and Genomes
miRNA	Micro Ribonucleic Acid
MSR	Master Sex Regulation
NF- κ B	Nuclear Factor of Kappa light polypeptide gene enhancer in B cells
PCA	Principle Components Analysis
PSD	Polygenic Sex Determination
qPCR	Quantitative Polymerase Chain Reaction

RAPD	Random Amplification of Polymorphic DNA
SD	Sex Determination
SOM	Self-Organizing Map

Abstract

Zebrafish is a popular research organism in many areas of developmental biology. However, knowledge on its sexual development is lacking. Before this project there were conflicting reports on zebrafish sex determination (SD) mechanism. There were reports that suggested zebrafish SD depended on environmental factors, while others claimed the presence of sex chromosomes. In addition to the confusion on SD mechanism, there was also a lack of information on when the bipotential gonad starts the testis or ovary differentiation process. This Ph.D. project aimed to address these two areas of zebrafish sexual development by using a combination of classical breeding and molecular methods.

The first part of this project sought to unravel the SD mechanism of zebrafish. Using classical breeding experiments, it was observed that progeny sex ratio across 62 zebrafish families were wide ranging (4.8% to 97.3% male) and included several sex-biased families. On the other hand, broods that were produced by repeated crossing of the same breeding pair had very similar sex ratios indicating strong influence from parental genotypes. In the selective breeding experiment, a sex ratio shift was observed and an all-male family was produced in four generations. These responses to selective breeding indicate that zebrafish SD mechanism is genetically based. To find out whether there is any substantial structural difference between male and female genomes, an indication for the presence of sex chromosomes, two molecular methods (namely FluoMEP and array comparative genomic hybridization) were developed and customized, respectively, for use in zebrafish. Extensive comparisons between genomes of the two sexes did not identify any universal structural difference. The lack of

substantial sex-linked difference between male and female genomes and the presence of sex-biased families are both incompatible with the chromosomal sex determination system. Together with the observations of wide ranging family sex ratios, strong parental genotype influence and responses to selective breeding show that zebrafish primarily uses a polygenic sex determination system.

After establishing the SD mechanism of zebrafish, the second part of this project was to attempt to specify the SD timing by finding the earliest sign of gonadal transcriptome bifurcation. Using *Tg(vasa:vasa-EGFP)* transgenic line, the gonadal transcriptome of zebrafish that showed strong (mostly future female) or weak (future male) vasa-EGFP fluorescence at 28, 35, 42 and 105 dpf were analyzed by a custom-made expression microarray. A total of 16,524 transcripts were found to be differentially expressed between the strong or weak fluorescence groups at one (or more) time point(s). Eleven of the up-regulated transcripts in the strong fluorescence groups from 28 to 105 dpf and nine candidate sex genes were used to make a qPCR array for analyzing zebrafish larvae (7, 14 and 21 dpf). Based on the expression profiles of the twenty selected genes, zebrafish larvae could be clustered into at least two groups from 7 dpf onward.

To summarize, the data obtained through the use of classical breeding experiments and customized genomic screening tools clearly indicate that the primary SD mechanism of zebrafish is a polygenic sex determination (PSD) system. From gonadal transcriptome profiling using a selected set of genes, divergent expression patterns were observed in zebrafish larvae as young as 7 dpf. This shows that gonadal differentiation begins prior to juvenile ovary phase. Therefore, we

hypothesize that zebrafish sex is determined between the period of fertilization to 7 dpf, before any sexually dimorphic histological phenotypes can be observed.

1. Introduction

1.1 Zebrafish: A teleost model organism

Teleosts (ray-finned fishes) form the largest group of extant vertebrates with about 32,000 species identified to date (Nelson 2007; Froese and Pauly 2013). It is believed that teleosts diverged from tetrapods around 420 million years ago (Chen *et al.* 2012) and had undergone an additional round of genome duplication compared to other animal groups (Meyer and Van de Peer 2005) resulting in rich diversity at both species and gene levels (Ravi and Venkatesh 2008). Rich diversity is also observed for teleost sexual developmental processes [reviews: (Ohno 1999; Mank and Avise 2009)] as all currently known sex determination (SD) mechanisms are represented in this group [review: (Devlin and Nagahama 2002)]. This unique feature of teleosts allows for comparative studies on the evolution of SD mechanisms.

One of the most popular model organisms among teleosts is zebrafish (*Danio rerio*). It is used in many areas of research from studying the biological effect of environment pollutants [review: (Segner 2009)] to deciphering molecular pathways [e.g. apoptosis; review: (Eimon and Ashkenazi 2010)] and recently gaining popularity as a model for human diseases research [review: (Lieschke and Currie 2007)]. Among the reasons that make zebrafish a popular laboratory model is its short generation time. In the laboratory, zebrafish reaches reproductive maturity at around 3 to 4 months. Despite its relatively small size, the species is

quite fecund; mature females kept under ideal conditions often produce 400-500 eggs regularly on a weekly basis.

Natural habitats of zebrafish stretch from South Asia (e.g. Pakistan, Nepal and India) to Southeast Asia such as Myanmar (Spence *et al.* 2008). This small-bodied freshwater species is most commonly found in slow or stagnant waters, such as rivers, ponds and paddy fields (Engeszer *et al.* 2007; Spence *et al.* 2008). The ambient water temperature at these natural habitats typically ranges between 26 to 32 °C (Engeszer *et al.* 2007).

1.2 Sex determination

Sexual reproduction is the most common method of reproduction in eukaryotes which involves random combination of gametes and recombination of genetic materials originating from the two parents. These two features are essential in maintaining genetic diversity of the population for adaption. Two of the most fundamental processes in reproductive biology are SD followed by gonad differentiation.

Sex determination is the process which decides the sexual fate of an organism. This biological event has profound effects on future secondary sexual developments, such as puberty and reproduction. Vertebrate SD mechanisms can be divided into two general groups, which are either regulated by genetic or environmental signals [reviews: (Devlin and Nagahama 2002; Manolakou *et al.* 2006; Angelopoulou *et al.* 2012)].

1.3 Genetic (genotypic) sex determination

In genetic/genotypic sex determination (GSD), the sex of an individual is determined by genes or chromosomes inherited from the parents. Therefore, for this mode of SD system sex is determined during conception. The most common type of GSD system utilized by mammals and avians is chromosomal sex determination (CSD) [reviews: (McLaren 1988; Manolakou *et al.* 2006; Ayers *et al.* 2013)]. Since most major vertebrate model organisms studied in detail so far seem to follow the CSD model, the vast majority of our knowledge on molecular regulation of vertebrate sex was collected from such systems [review: (Manolakou *et al.* 2006)]. CSD can be further sub-divided into male heterogametic (XX: female; XY: male) or female heterogametic (ZW: female; ZZ: male) systems. In CSD, a master sex regulation (MSR) gene typically resides on a preferential chromosome, such as *SRY* on the mammalian Y chromosome (Sinclair *et al.* 1990) and *xdm-w* on the *Xenopus laevis* W chromosome (Yoshimoto *et al.* 2008). The sex of an individual offspring will be dictated by the presence or absence of the MSR gene present on the sex chromosomes it inherits. CSD is also utilized by some teleost species with novel MSR genes such as *Dmrt1bY* in *Oryzias latipes* (Matsuda *et al.* 2002), *Gsd^Y* in *Oryzias luzonensis* (Myosho *et al.* 2012), *Amhy* in *Odontesthes hatcheri* (Hattori *et al.* 2012) and *sdY* in *Oncorhynchus mykiss* (Yano *et al.* 2012) or a gene variant like the *Amhr2* with a missense SNP in *Takifugu rubripes* males only (Kamiya *et al.* 2012). However the MSR genes of most teleosts CSD species are still unknown [e.g. threespine stickleback (Peichel *et al.* 2004) and Nile tilapia (Eshel *et al.* 2012)]. To add to the diversity, some teleosts using primarily CSD system can be overridden by environmental factors such as

high temperature induced masculinization of Nile tilapia [(Abucay *et al.* 1999); review:(Baroiller *et al.* 2009)].

Other types of GSD are dosage-dependent sex determination and polygenic sex determination (PSD). Dosage-dependent sex determination system is the common mode of SD for invertebrate species [review: (Manolakou *et al.* 2006)] where sex is determined by the ratio of X chromosome to autosomes [review: (Cline 1993)]. In the fruit fly (*Drosophila melanogaster*), individuals with a single X chromosome (XO) will become males while those which have two X chromosomes (XX) develop as females (Bridges 1921). In contrast, sex of PSD species is determined by cumulative effect of multiple genes or loci distributed among the autosomes [(Kosswig 1964); review: (Moore and Roberts 2013)]. Characteristics of a PSD system are 1) strong influence from parental genotypes on the offspring sex ratio; 2) huge variation in sex ratio among families; and 3) sex ratio shift in response to selection (Bull 1983). Due to the complexity of PSD and limited technologies for studying polygenic trait, there are limited published data on this system. Currently only a few fish species [e.g. European seabass (Vandeputte *et al.* 2007), Florida red tilapia (Desprez *et al.* 2006) and several Lake Malawi cichlid species (Ser *et al.* 2010)] are known to utilize the PSD system.

1.4 Contradicting data on the zebrafish SD mechanism

1.4.1 Cytogenetic methods to search for sex chromosomes

To find out if a species' primary SD mechanism is sex chromosome-based, the most direct method is to search for sex chromosomes in the genome. The classical

method for identification of sex chromosomes is detection of a size heteromorphic chromosomal pair in one of the two sexes by karyotyping. Highly differentiated sex chromosomes (e.g. most mammalian ones) are easily distinguished cytogenetically due to their distinct size difference (heteromorphism) caused by suppression of recombination and degeneration of the heterogametic sex chromosome [reviews: (Graves 1995; Delbridge and Graves 1999; Ellegren 2011)]. Search for a heteromorphic chromosomal pair in zebrafish started nearly half a century ago when its karyotype was described (Post 1964). Since then there were a number of publications on zebrafish karyotypes with majority of them reporting absence of a heteromorphic chromosomal pair (Post 1964; Endo and Ingalls 1968; Rishi 1976; Schreeb *et al.* 1993; Pijnacker and Ferwerda 1995; Daga *et al.* 1996; Gornung *et al.* 1997; Amores and Postlethwait 1998; Ueda and Naoi 1999; Phillips and Reed 2000), while two saw potential indications of a female heterogametic (ZZ/ZW sex chromosomes) system (Fontana *et al.* 1970; Sharma *et al.* 1998). The conflicting karyotype results arises from difficulties to pair zebrafish chromosomes due to their monomorphic nature and inconsistent chromosome banding patterns caused by usage of different chromosome stains in the various studies (Amores and Postlethwait 1998), thus making cross-validation of the results from different experiments challenging. Also, teleosts with cytogenetically recognizable sex chromosomes seem to be very rare as two independent studies which together reviewed ca. 3,000 karyotyping results (Arkhipchuk 1995; de Almeida Toledo and Foresti 2001) found less than 5% of them have heteromorphic sex chromosomes. Indeed, among teleost model organisms that primarily use CSD system, many of them do not have

heteromorphic sex chromosomes [e.g. Japanese medaka (Uwa and Ojima 1981) and threespine stickleback (Chen and Reisman 1970)].

Other cytogenetic methods for detecting sex chromosomes (not based on size dimorphism) are comparative analysis of meiotic recombination rates in males vs. females and synaptonemal complex karyotyping. Species with heteromorphic chromosomal pairs usually display different recombination rates that are sex-specific due to suppression of recombination between the pair of sex chromosomes (Tease and Hulten 2004). For example, teleosts with male heterogametic sex system [e.g. Japanese medaka (Matsuda *et al.* 1999) and rainbow trout (Danzmann *et al.* 2005)] were found to have lower recombination rate in the male (XY) when compared to the female (XX). For zebrafish, Postlethwait's group compared the recombination rate between a double-haploid male (Singer *et al.* 2002) and female (Postlethwait *et al.* 1994; Postlethwait *et al.* 1998) using genetic mapping. They reported that the male androgenetic map had lower recombination rate relative to the female gynogenetic map (Singer *et al.* 2002). Later, they also observed similar results in wild-type zebrafish (Anderson *et al.* 2012). However, in another similar study it was suggested that the lower recombination rate observed in male zebrafish is most likely due to lesser numbers of *human mutL homolog 1 (mlh1)* foci in the male genome and not because of recombination suppression (Kochakpour and Moens 2008). The latter observation was supported by earlier data obtained through synaptonemal complex karyotyping (Traut and Winking 2001; Wallace and Wallace 2003), which did not find any differences between the two zebrafish sexes. The contradicting result from these two sets of data shows that cytogenetic methods are unable to resolve with confidence whether sex

chromosomes are present in zebrafish or not. If sex chromosomes are present, then it is unlikely that they are highly differentiated.

1.4.2 Detection of the CSD system by whole genome manipulation

Since the absence of cytogenetically recognizable heteromorphic chromosomal pair does not rule out the presence of sex chromosomes, the search was brought to genomic level. Whole genome manipulation (or induced parthenogenesis) techniques can be used to identify CSD species by analyzing sex ratio of the F_1 offspring or subsequent generation's (F_2) progeny sex ratio [reviews: (Purdom 1969; Parsons and Thorgaard 1985; Ihssen *et al.* 1990; Horváth and Orbán 1995; Komen and Thorgaard 2007)]. The principle of whole genome manipulation is based on the exclusion of paternal or maternal genome from inheritance by irradiation of sperm or egg, followed by inhibition of first cell division. This will result in duplication of the remaining haploid genome in the embryo, hence the name double haploids (DH) [review: (Komen and Thorgaard 2007)]. DHs generated by the exclusion of paternal genome are called gynogenotes, whereas those produced by the removal of maternal chromosomal set are called androgenotes. If the species uses CSD system, the DHs generated will have equal chances to inherit either one of the two sex chromosomes. For species that are male heterogametic (XY), the expected sex ratio of F_1 DH from gynogenesis would be 100% female (Figure 1A). On the other hand, for female heterogametic species (ZW) their F_1 gynogenotes are expected to have a 1:1 male to female phenotypic sex ratio (Figure 1B). All the female gynogenotes should have a WW sex chromosomal set that does not occur in nature. If the WW female is viable and crossed with a normal male (ZZ) they should produce an all-female progeny

(Figure 1B). The reverse is true for androgenesis whereby F₁ DHs from male heterogametic species should have an equal proportion of males and females, while female heterogametic species should produce 100% male androgenotes. For zebrafish, sex ratio data obtained from four sets of gynogenesis (Streisinger *et al.* 1981; Hörstgen-Schwark 1993; Pelegri and Schulte-Merker 1999; Uchida *et al.* 2002) and one androgenesis experiment (Corley-Smith *et al.* 1996) yielded contradictory indications (for summary of data see Table 1). The number of DHs obtained per experiment ranged from five to 146 and sex ratios showed high variation from 0% to 95% females. Some of these data suggested either a male or a female heterogametic system, while the rest did not seem to fit either of the two systems. Again the data is inconclusive about the presence or absence of sex chromosomes, but taken together they seem to argue against CSD as the primary SD mechanism of zebrafish.

Figure 1: Inferring the type of sex chromosomal system by progeny sex ratios as a result of genome manipulation or artificial sex reversal.

(A) Performing gynogenesis on male heterogametic (female homogametic) species will produce gametes with only X chromosome, therefore the F_1 progeny sex ratio will be close to 100% female. (B) The female heterogametic species gynogenesis will produce gametes with either Z or W chromosome. The resulting F_1 phenotypic sex ratio will have equal proportion of male and female with all F_1 females having WW sex chromosomes (assuming full viability of all genotypes). When F_1 gynogenotes females (WW) are crossed with normal males (ZZ), their F_2 offspring will be all female (ZW). (C) Using hormone- or temperature-based masculinization on male heterogametic species will produce sex-reversed neo-males (srXX) and when crossed with normal females will produce all-female F_1 offspring. (D) The sex-reversed neo-male (srZW) of female heterogametic species will have both sex chromosomes. When the sex-reversed neo-male (srZW) is crossed with normal female (ZW), the expected phenotypic sex ratio in the F_1 progeny will be 25% males and 75% females. Among the F_1 females, one-third of them will have WW sex chromosomal pair that does not normally occur in nature and yield an all-female offspring (F_2 ; not shown) when crossed with a normal male (ZZ). Upper panel (A & B): genome manipulation; lower panel (C & D): artificial sex reversal.

Table 1: Published data from induced parthenogenesis experiments on zebrafish yield sex ratios incompatible with one or both CSD systems.

Double Haploid (DH)	Strain	Number of DH	Obs sex ratio (% ♀)	Type of CSD system excluded	Remark	Reference
Gynogenotes ¹	C29	85	95%	ZW/ZZ	If XX/XY, then YY viability is high	Streisinger <i>et al.</i> , 1981
Gynogenotes ¹	Wild	8	0%	XX/XY	If ZW/ZZ, then WWs are inviable	Hörstgen-Schwark, 1993
Gynogenotes ¹	Golden	8	0%	XX/XY	If ZW/ZZ, then WWs are inviable	Hörstgen-Schwark, 1993
Androgenotes ²	AB	5	0%	both	Sex ratio of F ₂ progeny is highly skewed in some families	Corley-Smith <i>et al.</i> , 1999
Gynogenotes ¹	gol-mix	15	<20%	both	Per family sex ratios cover the range from <20% to >80%	Pelegri & Schulte-Merker, 1999
Gynogenotes ¹	gol-mix	146	<20%	both	Per family sex ratios cover the range from <20% to >80%	Pelegri & Schulte-Merker, 1999
Gynogenotes ¹	gol-FL1	28	92%	both	Per family sex ratios cover the range from <20% to >80%	Pelegri & Schulte-Merker, 1999
Gynogenotes ¹	Unknown	Unknown	6%	XX/XY	If ZW/ZZ, the viability of WWs is very low	Uchida <i>et al.</i> , 2002

1: Expected F₁ sex ratio for gynogenotes of male heterogametic species is 100% female and female heterogametic species is 50% female.

2: Expected F₁ sex ratio for androgenotes of male heterogametic species is 50% female and female heterogametic species is 0% female.

1.4.3 Breeding sex-reversed fish to resolve the SD mechanism

Another experimental technique used for detecting the presence of sex chromosomes is based on analyzing progeny sex ratios produced by crossing sex-reversed individuals with wild-type partners [review: (Pandian and Sheela 1995; Devlin and Nagahama 2002; Piferrer *et al.* 2012)]. In male heterogametic species, when sex-reversed neo-males (srXX) produced by masculinization of genetic females are crossed with wild-type females, a monosex female progeny is expected (Figure 1C). However, in female heterogametic species, the proportion of females in the progeny of neo-male (srZW) cross with normal female (ZW) is expected to range from 50% to 75% (Figure 1D), depending on the relative survival rate of the WW individuals compared to ZW. When neo-females generated by feminization of genetic males are crossed with wild-type males, the expected sex ratios are: 25-33% females for male heterogametic species (depending on the survival of the YYs) and 100% males for female heterogametic species.

Sex reversal of zebrafish by hormonal treatment was reported for numerous experiments [for a set of examples, see Table 1 of (Orban *et al.* 2009)]. However, according to our knowledge, most of them did not report the progeny sex ratios produced by sex-reversed neo-males or females. We only managed to find one study that successfully masculinized zebrafish by sex hormone treatment and identified ‘super females’ among the resulting F₁ fish (Tong *et al.* 2010). When these ‘super females’ were crossed with normal males severely female-biased (90-98%) progenies were produced. This led them to suggest that zebrafish has a ‘female dominance chromosomal based sex determination mechanism’. This proposed mechanism is similar to the female heterogametic system (ZW/ZZ) that

was found to be incompatible with zebrafish SD based on data obtained earlier by whole genome manipulation (Table 1). Unfortunately, there is no indication in the publication whether the authors have attempted to identify the proposed sex chromosomes (or any of their markers).

Masculinization of zebrafish can also be achieved by high temperature applied during the gonad differentiation window [(Uchida *et al.* 2004; Abozaid *et al.* 2011, 2012); review: (Ospina-Alvarez and Piferrer 2008)]. In an experiment conducted by the Hörstgen-Schwark's group (Abozaid *et al.* 2012), F₁ neo-males produced by heat treatment were backcrossed to their mother (normal female) for progeny sex ratios analysis. On average the F₁ progenies had 16% males, which was a notable difference from the expected sex ratio for both male heterogametic (expected: no males) and female heterogametic (expected: 50% males) systems. Surprisingly, the F₂ progenies had sex ratios very similar to the expected value of a female heterogametic system. However, since the progeny sex ratios from the F₁ generation could not be explained by either of the two CSD systems, the authors concluded that at least two or more autosomal factors might be involved in zebrafish SD, further strengthening indications against the compatibility of the system with a simple sex chromosomal model.

1.5 Environmental factors exert secondary effects on zebrafish SD

The other major type of SD mechanism is environmental sex determination (ESD). As indicated by its name, sexual fate of an ESD species is determined by signal from its surrounding environment (Charnov and Bull 1977). Unlike GSD species whose sex is determined at the point of fertilization, the sex of ESD species is only

determined after fertilization during a crucial embryonic developmental stage (Charnov and Bull 1977). Another difference between GSD and ESD species is that the latter do not have sexual dimorphism at genomic level.

There are also various types of ESD system and the most studied type is temperature-dependent sex determination, in which sex is a function of the ambient temperature under natural condition during early development. This mode of SD is prevalently used by reptiles (Pieau *et al.* 1999; Modi and Crews 2005) and also exists in some teleost species (Ospina-Alvarez and Piferrer 2008). However, it was also observed that many teleost species with GSD system respond to heat-induced masculinization too (Ospina-Alvarez and Piferrer 2008).

Studies on various teleost species, including zebrafish, showed that elevated water temperature affected different gonadal processes, including apoptosis, estrogen biosynthesis and germ cell meiosis (Uchida *et al.* 2004; Navarro-Martin *et al.* 2011; Yamaguchi and Kitano 2012). When juvenile zebrafish were subjected to elevated water temperatures (36 °C) during the gonad transformation period [ca. 18-28 day post fertilization (dpf)], oocyte apoptosis and decreased activity of gonadal aromatase was observed resulting in increased proportion of males at adulthood (Uchida *et al.* 2004). In European seabass (*Dicentrarchus labrax*) high temperature caused increased DNA methylation of the aromatase promoter resulting in decreased expression of *cyp19a* and male-biased sex ratio (Navarro-Martin *et al.* 2011). Apparently, the masculinizing effect of high temperature is not restricted to gonad differentiation period in zebrafish, as families exposed to increased temperature during larval stage [5-42 hours post fertilization] also

produced significantly more males (Abozaid *et al.* 2011). Interestingly, in the latter experiment some families tended to be more sensitive to heat-induced masculinization than others (Abozaid *et al.* 2012). This shows that interaction between genotype and environment exists in zebrafish.

In addition to temperature, several other environmental factors were also found to have an effect on zebrafish sex ratio. Concentration of dissolved oxygen in the water is one of the environmental factors. Zebrafish kept at low oxygen (hypoxia) condition (0.8 mg O₂/L) from 0 to 120 dpf showed a significant increase of males (difference of 12.5% male) compared to normoxia (5.8 mg O₂/L) condition (Shang *et al.* 2006). Hypoxia is believed to affect expression of genes in the steroidogenic and serotonergic pathways resulting in decreased female sex hormone biosynthesis (Shang *et al.* 2006; Martinovic *et al.* 2009). Growth rate is another environmental factor reported to have an effect on zebrafish sex ratio (Lawrence *et al.* 2008). It was observed that zebrafish broods that grew faster by frequent feeding tend to have more females. This led researcher to hypothesize that growth rate indirectly regulates sex differentiation, because growth is linked to the regulation of sex hormone biosynthesis. On the other hand, a recent study found that zebrafish females will only reach puberty if they exceeded a threshold weight (100 mg) and length (1.8 cm) (Chen and Ge 2013). They observed that female zebrafish which were over the threshold were fertile as early as 60 dpf, while those below the threshold were infertile even at 90 dpf. Therefore, the group receiving less food in the study by Lawrence and co-workers (Lawrence *et al.* 2008) could have contained immature females that were unknowingly sexed as male. Furthermore, rearing density which is an environmental factor closely related to growth rate, was

found to have no influence on progeny sex ratio by two independent studies covering zebrafish from larvae to juvenile stages (Hazlerigg *et al.* 2012; Liew *et al.* 2012).

All in all, the above studies showed that zebrafish sexual development could be affected by various environmental factors. However, many of the environmental factors do not exert a strong effect and some of them seemed to be largely dependent on the fish's genotype. These data argue against zebrafish being a true ESD species and further support the earlier conclusion of Piferrer's group that zebrafish sexual development merely responds to high temperature (Ospina-Alvarez and Piferrer 2008).

1.6 Zebrafish gonads go through a mandatory “juvenile ovary” phase

The commitment to male or female fate by sex determination is followed closely by gonad (sex) differentiation. This process initiates major restructuring of the bipotential gonad into either testis or ovary with involvement of multiple gene networks [review: (Piferrer and Guiguen 2008; Orban *et al.* 2009)].

The complexity of zebrafish gonad differentiation process was first reported by Takahashi (Takahashi 1977), who pointed out that every individual forms an ovary first (later it was named ‘juvenile ovary’). Gonad of future males will then have to undergo a transformation prior to testis differentiation. The ‘juvenile ovary’ is made up of mostly oogonia and primary oocytes (Maack and Segner 2003). First

histological signs of the ‘juvenile ovary’ phase can be observed from zebrafish as young as 2.5 to around 4 weeks post fertilization (Takahashi 1977; Maack and Segner 2003). Extent of the ‘juvenile ovary’ phase varies among individuals from a few days to more than a week (Wang *et al.* 2007). Following ‘juvenile-ovary’ phase, the gonad either continues to develop into a functional ovary or undergoes a drastic transformation to form a testis [review: (Orban *et al.* 2009)].

The hallmark of this ‘juvenile ovary to testis’ transformation process is a rapid degeneration and eventual elimination of oocytes by an apoptotic (or programmed cell death) wave that was first described by Uchida and colleagues (Uchida *et al.* 2002). Although the primary signal that triggers the apoptotic wave is still unknown, it is likely that it acts through a change of hormonal environment around the primary oocytes that result in increased ratios of male to female hormones. The numbers of primordial germ cells (PGC) also seem to have a role in the transformation process as their complete depletion leads to the development of sterile males (Slanchev *et al.* 2005; Siegfried and Nusslein-Volhard 2008). In the adult, germ cells are also required for maintenance of ovarian function (Dranow *et al.* 2013).

1.7 Multiple genetic pathways are involved in gonad differentiation

The improvement of technologies, especially the broadening application of expression microarrays and characterization of mutants with relevant phenotypes [e.g. *fanc1*^{*nkhg10aEt/nkhg10aEt*} (Rodríguez-Marí *et al.* 2010)] has allowed researchers to start looking for large-scale changes and study the involvement of major developmental pathways in gonad differentiation. Through analysis of the

zebrafish *fanc1*^{nkhg10aEt/nkhg10aEt} mutant line with mutation in *Fanconi Anemia complementation group L* (*fanc1*; NM_212982) gene, the Postlethwait's group had obtained new information on the regulation of apoptotic processes during gonad transformation (Rodríguez-Marí and Postlethwait 2011). According to their results, zebrafish *tp53* [an ortholog of the human tumor protein *P53*; (Lane and Crawford 1979)] activates the apoptotic processes in the gonad of males resulting in the removal of oocytes and thereby tilting the balance towards testicular differentiation (Rodríguez-Marí *et al.* 2010). Similar studies using mutants with mutation in other sex candidate genes can be used to unravel the complex network of developmental pathways involved in gonad development. A mutant line with such a potential is the *nanos3*^{fh49/fh49} line (Beer and Draper 2013). The zebrafish *nanos* family of genes was found to be essential for the maintenance and specification of female germline stem cells, which in turn regulate oocyte production (Draper *et al.* 2007; Beer and Draper 2013). Using mutant lines with mutation in a gene having such important function for transcriptomic analysis will yield important information on ovary development.

The involvement of NF-κB [nuclear factor of kappa light polypeptide gene enhancer in B cells; (Hayden and Ghosh 2008)] pathway in zebrafish gonad differentiation was also discovered recently. Juvenile zebrafish subjected to heat-killed bacteria showed a drastic change of adult sex ratio towards female bias (Pradhan *et al.* 2012). Transcriptomic analysis of the underlying processes identified up-regulation of NF-κB pathway by the induced inflammation as a cause (Pradhan *et al.* 2012). When the pathway was chemically inhibited the sex ratio shifted towards male bias (Pradhan *et al.* 2012).

Our group has also performed a microarray-based transcriptomic analysis of the developing juvenile gonad (Sreenivasan *et al.* 2009) using the *Tg(vasa:vasa-EGFP)* line developed by the Olsen lab (Krøvel and Olsen 2002). The analysis of differentially expressed genes indicated potential involvement of the canonical Wnt (or Wnt/beta-catenin) signaling pathway (Sreenivasan *et al.* 2009) that was shown earlier to play an important role in mammalian ovary formation and maintenance [reviews: (Bernard and Harley 2007; Nef and Vassalli 2009)]. Subsequent down-regulation of the pathway through transgenic inhibition has resulted in a significant decrease of the proportion of males, providing functional proof for its importance in ovary development and possibly maintenance (Sreenivasan *et al.* 2014).

In summary, currently one pro-male (Tp53-activated apoptosis) and two pro-female (NF- κ B and canonical Wnt) pathways were found to be involved with gonad transformation and differentiation in zebrafish. All of them have an effect on apoptotic processes that control the fate of cell types in developing gonads. The shift of balance of these three pathways into one or the other direction will push the development towards one of the two gonad types (Figure 2). Studies using mammalian model organisms showed that other pathways, such as retinoic acid pathway (Bowles *et al.* 2006; Koubova *et al.* 2006) and Fgf9 pathway (Bowles *et al.* 2010), were also involved in gonad differentiation. Therefore it seems likely that additional pathways will have a role in zebrafish gonad differentiation and further studies will be needed to unravel all the details of this complex process.

Figure 2: Shift in the balance of pro-male and pro-female pathways will determine the direction of gonad differentiation in zebrafish.

Three major developmental pathways (Tp53, NF-κB and canonical Wnt pathways) have been shown to participate in gonad differentiation process. In males (left), the numbers of primary oocytes originating from the ‘juvenile ovary’ are low, Tp53 signaling pathway is up-regulated, apoptotic processes are fully active while the NF-κB as well as canonical Wnt signaling pathways are both down-regulated. In females (right), the number of primary oocytes originating from the ‘juvenile ovary’ is high, NF-κB and canonical Wnt signaling pathways are both up-regulated, whereas Tp53 signaling pathway is down-regulated. The result is suppression of apoptosis. Currently the timing and causative effects of these processes are not fully understood.

1.8 Objectives of this study

Despite being a model research organism in the field of developmental biology, essential knowledge about zebrafish sexual development is still lacking. Various cytogenetic and genome manipulation studies spanning over four decades have been unable to convincingly identify the SD mechanism of zebrafish. For gonad differentiation, various publications had provided more knowledge about the process; however they had also revealed that it is more complicated than originally

thought. For instance, it is still not known when the gonad transformation process starts and if more biological pathways are involved.

Hence the objectives of this study were:

First, to elucidate the SD mechanism of zebrafish by analyzing a large number of family sex ratio data collected over a decade by previous lab members and during this study.

Second, to use molecular techniques (i.e. FluoMEP and aCGH) to decide whether sex chromosomes are present in the zebrafish genome or not.

Third, to design an expression microarray based on the latest zebrafish genome assembly (Zv9) for studying gonad transformation.

Fourth, to use expression profiling of genes with sexually dimorphic expression pattern to specify the time period of SD in zebrafish.

The overall objective of this project was to advance the field of zebrafish sexual development research. New knowledge obtained from this study can be extrapolated to research on other important aquaculture species and potentially even to that of mammalian sex development disorders.

2 Materials and Methods

2.1 Fish stocks and tail fin samples

All experiments with live fish were approved by Temasek Life Sciences Laboratory Institutional Animal Care and Use Committee (approval ID: TLL (F)-10-001). A wild-type zebrafish (*Danio rerio*) strain, called Toh, used in this study was bought from a local farm. The Tuebingen (Tu) zebrafish strain was a kind gift from Dr Matthew Harries (Max Planck Institute for Developmental Biology, Tuebingen, Germany). The *Tg(vasa:vasa-EGFP)* zebrafish transgenic line (AB strain based) was a kind gift from Dr Lisbeth Olsen (Sars International Centre for Marine Molecular Biology, Bergen, Norway). All zebrafish were kept in AHAB (Aquatic Habitats) recirculation systems according to standard protocols (Westerfield 2000). Guppy (*Poecilia reticulata*) fin clip samples were kind gifts from Prof Rob Brooks (University of New South Wales, Sydney, Australia). Rosy barb (*Puntius conchoni*) fish were purchased from a local fish trading company (Qian Hu Fish Farm, Singapore) and tail fin samples were collected under anesthesia then stored in absolute ethanol at -20 °C until use.

2.2 Fish husbandry

Adult zebrafish were kept as mixed sex groups in 2.75 L tanks at a density of <10 fishes per liter. Breeding was carried out in meshed-bottom mouse cages of one liter volume placed into a second cage containing egg water. Breeding pairs were set up at the previous evening in the presence of artificial plants and eggs were collected before noon the next day. Pairs that were reluctant to yield eggs were given a slight cold shock by replacing about 25% of the water with fridge cooled

(ca. 4 °C) water one to two hours after the start of breeding period. Ripe females that failed to produce eggs after two breeding attempts with different males were gently squeezed at the abdominal area. This was to aid the removal of eggs, if any, that might be ‘stuck’ in the oviduct then set up for repeated mating after a one week rest.

Fertilized eggs were collected from the bottom of the cage, rinsed on a tea filter and transferred into plastic trays with egg water containing methylene blue. Survivals were recorded at one and two dpf. Embryos were transferred onto the AHAB system after hatching and they were grown there at the following densities (unless indicated otherwise): <100/L for embryos, <80/L for larvae, <20/L for juveniles and <10/L for young adults.

2.3 Visual sexing of zebrafish

Adult zebrafish were sexed visually, based on the following two criteria: i) general body shape; and ii) the presence of ‘anal papilla’ in females (observed on unstressed fish kept in water). Individuals with intermediate body shape and poorly observable papilla were gently squeezed and checked for eggs or sperm. In the absence of either, individuals were euthanized, dissected under stereo-microscope (Leica) and their gonad analyzed by gonadal squash. Those individuals with unclear sex were not included in the sex ratios study.

2.4 Selective breeding experiment

We had carried out an experiment to tilt families toward male- or female-biased sex ratio by performing up to four generations of selective breeding. Over 60 crosses were performed using sexually mature adults from our own stocks (mostly AB and Toh strains). Based on the progeny sex ratios we created two lines to be used for selection against males or females. Typically, sexually matured full-sib males and females (three of each sex) were selected based on their phenotype and robustness. Pairwise full-sib crosses were performed according to a multifactorial design. The offspring were sexed at about 3 months of age and family sex ratio recorded. Three robust males and three robust females were then chosen from crosses that produced highly skewed sex ratio as brooders to produce the next generation.

2.5 FluoMEP assay

Fluorescent motif enhanced polymorphism (FluoMEP), a new PCR-based assay for genomic screening (Figure 3) was developed during this project and carried out as described previously (Chang *et al.* 2007). First, genomic DNA (gDNA) samples were extracted from tail fins by incubating them at 55 °C overnight in 800 µl of SET buffer [0.5% (w/v) SDS, 50 mM EDTA, 10 mM Tris/Cl (pH 8), 200 mM NaCl] and 250 µg/ml of Proteinase K (Roche). Then standard phenol chloroform extraction (Sambrook and Russell 2006) was performed and the gDNA pellet was dissolved in 100 µL of TE buffer. Two pools of samples were generated by combining individual gDNA samples (9 individuals of each sex for rosy barb, 8 individuals of each sex for guppy and 4 individuals of each sex for zebrafish).

After that the pooled male and female gDNA samples were compared by bulk segregant analysis (Michelmore *et al.* 1991) to identify potential sex-linked markers. These potential markers were then used on individual samples for confirmation. See Appendix 1 for primer sequences of guppy and rosy barb male-specific FluoMEP markers.

The FluoMEP assay was carried out as described previously (Chang *et al.* 2007). Briefly, the PCR mixture consisted of 15 pmol of fluorescently labeled common primer, 7.5 pmol of random amplification of polymorphic DNA (RAPD) primer (Eurofins MWG Operon), 0.2 μ M of dNTP, 2.5 mM of MgCl₂ and one unit of *Taq* (NEB). The FluoMEP PCR condition consisted of 40 cycles of 92 °C for 20 sec, 42 °C for 20 sec then 72 °C for 90 sec and finally a single step of 72 °C for 5 min.

2.6 Array comparative genomic hybridization

Genomic DNA samples for array comparative genomic hybridization (aCGH) were extracted using DNeasy Blood and Tissue kit (Qiagen) according to manufacturer's instructions with slight modification. Instead of 2 hours incubation, tail fin samples were incubated overnight at 55 °C in lysis buffer AL and proteinase K with slow shaking (70 rpm). The quality of extracted gDNA samples was checked on Nanodrop 1000 (Thermo Scientific) and by agarose gel electrophoresis. Samples that showed smear in the gel (sign of degradation) were not used and gDNA extraction was repeated. Each zebrafish family analysed consisted of the male parent (sire), female parent (dam), two male offspring and two female offspring.

Two aCGH microarrays were custom designed by Roche Nimblegen for this project. One had 108,417 probes (HX12) on the Nimblegen 12 x 135K format and the other had 661,018 probes (HX3) on the Nimblegen 3 x 720K format (see Appendix 2 for probes distribution on each chromosome).

Figure 3: Schematic diagram of the FluoMEP principle.

This PCR-base method uses a combination of commercial RAPD primers (blue arrows) with fluorescent labeled “common primer” (red arrows) which are designed to target common motif (e.g. LINE and SINE) in the genome (red rectangle). The amplified products consist of three kinds of DNA fragments: 1) those amplified by RAPD primers only (both ends blue); 2) by common primers only (both ends red); and 3) by combination of RAPD and common primers (one blue and one red end). High throughput capillary electrophoresis (e.g. ABI 3730xl DNA Analyzer) was used to analyze the amplified fragments. Only fragments amplified by at least one fluorescently labeled common primer (i.e. product 2 and 3) can be detected by the capillary electrophoresis machine. A software called FluoMEP Marker Finder (Sukumaran 2008) was developed in the lab to find polymorphic DNA marker from the huge amount of electrophoresis data.

Samples were labeled with NimbleGen Dual-Color DNA Labeling Kit (Roche NimbleGen) and hybridization was carried out according to the manufacturer's instructions with the MAUI hybridization system (BioMicro Systems). The oligo array was custom-designed by Roche NimbleGen based on the zebrafish Zv7 (danre5) genome assembly. During the course of this study two updates on the zebrafish genome assembly were released with Zv9 (danre9) being the latest version, all probes were re-mapped onto the new assembly for data analysis.

The array was scanned at 5 μ m resolution with an Axon GenePix 4000B Microarray Scanner (Molecular Devices). Raw fluorescent intensity data was retrieved by NimbleScan version 2.6 (Roche NimbleGen) then imported into Partek Genomic Suite version 6.6 (Partek Incorporated) for analysis. For variable copy number detection the genomic segmentation algorithm was used. A minimum of 3 markers were specified, p value threshold set at 0.001 and signal-to-noise ratio set at 0.3.

2.7 Validation of CNVRs by real-time quantitative PCR

Copy number variable regions (CNVRs) were validated by real-time quantitative PCR (gDNA-qPCR) as described in (Hoebeeck *et al.* 2007). Briefly, PCR primers targeting a sub-region of the CNVR were designed using Primer3 version 0.4.0 (Rozen and Skaletsky 1999). All samples were normalized with two reference loci (i.e. DrSC19 and DrSc23) found to be single copy exon in zebrafish genome (Li *et al.* 2007). Please refer to Appendix 3 for primer sequences. The gDNA-qPCR was performed with a Biorad MyIQ real-time PCR detection system (Biorad) and IQ

Sybr Green SuperMix (Biorad). Subsequently, relative copy numbers of each individual sample in the family was calculated with reference to the sire's (paternal) copy number according to the comparative Cq method described by Hoebeeck and colleagues (Hoebeeck *et al.* 2007).

2.8 RNA extraction

All samples used for RNA extraction were from offspring of the same breeding pair which produced progeny with adult sex ratio of 24% male. Trunk samples were dissected from euthanized *Tg(vasa:vasa-EGFP)* zebrafish larvae (7, 14, 21 and 28 dpf) between the opercula and anal pore. The trunk was gutted carefully using ultra-fine (1 µm tip) tungsten dissecting needle (Roboz Surgical Instrument Co.) under a stereo-microscope (Leica) equipped with MAA-03 universal light source (BLS Ltd) to ensure that the gonad (expressing vasa-EGFP) was still intact in the trunk cavity. For juvenile (35 and 42 dpf) and adult (105 dpf) zebrafish which were larger in size, gonadal samples were collected from euthanized fish. Dissection was carried out with guidance from the vasa-EGFP fluorescence signal which marked the gonad under stereo-microscope (Leica) equipped with MAA-03 universal light source. All tissue samples were immediately snap-frozen in liquid nitrogen then stored at -80 °C.

Total RNA were extracted using Ambion RNAqueous-Micro Kit (Life Technologies) except for the larger 105 dpf gonadal samples which were extracted using Ambion RNAqueous-4PCR Kit (Life Technologies). Contaminating gDNA was removed by treating with DNase I provided with the kit and according to the manufacturer's protocol. RNA quality was assessed by an Agilent 2100

Bioanalyzer with RNA 6000 Pico chip (Agilent Technologies). Only samples with RIN number ≥ 7.5 were used for subsequent microarray analysis. RNA concentration was quantified using a Qubit 1.0 Fluorometer with RNA Assay Kit (Life Technologies) according to manufacturer's instructions.

2.9 Expression microarray

The custom designed zebrafish expression microarray contained probes for 31,477 unique transcripts from a gonadal expressed sequence tag (EST) library created previously in the lab (Sreenivasan *et al.* 2008) and various public databases (Figure 4). Sequences from the various sources underwent a series of sequential clustering using USEARCH (Edgar 2010) with cutoff at 95% identity to remove redundant sequences. The microarray was manufactured by Roche NimbleGen on the 12 x 135K array format with each array containing 117,915 probes (length: 60 bp). On average, each transcript was targeted by three unique probes. During the course of this project there were several transcriptome updates therefore the probes were re-mapped to the most current Ensembl transcriptome version e70 (released on Jan 2013).

Samples used for the microarray experiment were from a single pairwise cross. They were collected starting from 28 dpf as that was the earliest developmental stage when vasa-EGFP fluorescence intensity could be visually sorted with acceptable efficiency under stereo-microscope. Three samples each from the weak and strong fluorescence groups at the four time points (28, 35, 42 and 105 dpf) were analyzed by microarray. Due to the small amount of RNA samples isolated from 28, 35 and 42 dpf gonadal samples (see section 2.8), whole transcriptome

amplification was necessary prior to microarray hybridization. 7.5 ng of total RNA was used for amplification with Ovation RNA Amplification System V2 (NuGEN). Size distribution of the amplified double stranded cDNA samples were checked by Agilent 2100 Bioanalyzer with RNA 6000 Nano chip (Agilent Technologies) to ensure fragment size uniformity among the samples. For the 105 dpf gonadal samples, 1 µg of total RNA was used to synthesize double stranded cDNA with cDNA Synthesis System kit (Roche) according to the protocol in Roche Gene Expression Technical Note (Skalitzky *et al.* 2011).

After cDNA synthesis, 1 µg of the double stranded cDNA from each sample was labeled using NimbleGen One-Color DNA Labeling Kit (Roche NimbleGen) and hybridization was carried out according to the manufacturer's instructions with the MAUI hybridization system (BioMicro Systems).

After overnight hybridization (16-20 hours) the microarray chip was washed with NimbleGen Wash Buffer Kit (Roche NimbleGen) according to manufacturer's protocol. The array was then scanned at 5 µm resolution with an Axon GenePix 4000B Microarray scanner (Molecular Devices).

Figure 4: Databases used to custom design Zv9 expression microarray.

Sequential clustering of sequences downloaded from various public databases (red boxes; accessed on Mar 2011) to remove redundant sequences. The zebrafish Gonadal Uniclone EST library was created previously in the lab (Sreenivasan *et al.* 2008). Numbers of sequences downloaded from each database are indicated in the box. The number of non-redundant sequences after each clustering was indicated under the arrow. During the course of this project the transcriptome was updated to version e70. The final data set contains 31,477 unique transcripts (genes and EST).

2.10 Expression microarray data analysis

The microarray data were collected and analyzed according to the Minimum Information About a Microarray Experiment (MIAME) guidelines (Brazma *et al.* 2001). Raw fluorescent intensity data were retrieved from the scanned images by NimbleScan version 2.6 (Roche NimbleGen) according to the manual. Robust multi-array average (Irizarry *et al.* 2003), quantile normalization and background

correction were applied as implemented in NimbleScan to generate the Pair files. The Pair files were then imported into Partek Genomic Suite version 6.6 (Partek Incorporated) for analysis. Differentially expressed transcripts (DETs) were identified by a 2-way mixed model ANOVA with factors vasa-EGFP fluorescence intensity and age. The p value was adjusted to account for incorrectly rejected null hypothesis in multiple hypotheses testing using False Discovery Rate (FDR) step-up method (Benjamini and Hochberg 1995) as implemented in Partek Genomic Suite. DETs between the strong and weak vasa-EGFP fluorescence groups for each time point (age) were identified by filtering for significant (FDR adjusted $p \leq 0.05$) then by at least 1.5 fold difference.

The microarray data was also used to identify gene sets that showed significant difference between the strong and weak fluorescence groups. This analysis was performed using GO-ANOVA as implemented in Partek Genomic Suite because it has the ability to analyze the entire microarray data as no pre-filtering is required. Hence, this approach is more powerful than other similar methods which require pre-filtering of the microarray data. Gene sets were obtained from gene ontology (GO) database (Ashburner *et al.* 2000) and Kyoto encyclopedia of genes and genomes (KEGG) pathways (Kanehisa and Goto 2000) for *Danio rerio*. Gene sets were filtered with average fold change of 1.2 and $p \leq 0.1$ (FDR corrected). Setting the p value at less than or equal to 0.1 was more stringent than the recommended 0.25 for other similar analytical methods (Subramanian *et al.* 2005).

2.11 High throughput real-time quantitative PCR array for expression analysis

A high throughput real-time qPCR array was designed to measure expression levels of: 1) 11 transcripts chosen randomly from the 69 DETs that were up-regulated in the strong fluorescence groups from 28 to 105 dpf in the microarray experiment; 2) nine candidate genes known to be involved in gonad development; 3) four randomly chosen transcripts for microarray validation; and 4) two endogenous reference genes. The primers were designed using online software Primer-BLAST (Ye *et al.* 2012) and their efficiency estimated with Real-time PCR Miner (Zhao and Fernald 2005). Please refer to Appendix 4 for genes details and primers information.

The qPCR array was performed using BioMark HD System and 96.96 Dynamic Array (Fluidigm). All qPCR reactions were performed in triplicates. Fifteen nanogram of total RNA from 7 (n=7), 14 (n=14), 21 (n=8), 28 (n=14), 35 (n=14) and 42 (n=14) dpf samples were used for first-strand cDNA synthesis using iScript Reverse Transcription Supermix (Biorad) according to manufacturer's manual and the reaction was incubated in PTC100 thermal cycler (MJ Research). Preparation of the BioMark Dynamic Array reactions was performed according to Fluidigm advance development protocol: Using EvaGreen DNA Binding Dye with 48.48 and 96.96 Dynamic Array IFCs (PN100-1208A1). Pre-amplification of cDNA was performed using TaqMan PreAmp Master Mix (Life Technologies) with 50 nM of pooled primers. Pre-amplification reaction was carried out at 95 °C for 10 min followed by 14 cycles of 95 °C for 15 sec and 60 °C for 4 min. The amplified cDNA was diluted 1:5 with low EDTA-TE buffer. For loading the BioMark

Dynamic Array, sample and assay mix was prepared according to manufacturer's protocol. Briefly the sample mix was prepared by combining 2 μ l of diluted amplified cDNA, 4.44 μ l of TaqMan Gene Expression Master Mix (Life Technologies), 0.443 μ l of 20x EvaGreen (Biotum), 0.443 μ l of 20x DNA Binding Dye Sample Loading Reagent (Fluidigm) and 0.67 μ l of low EDTA-TE buffer. The assay mix was prepared by combining 0.25 μ l each of the 100 μ M forward and reverse primer, 0.125 μ l of 10% (v/v) Tween 20 (Sigma) and 4.375 μ l of low EDTA-TE buffer. After the reaction mixture was loaded into the BioMark Dynamic Array, the qPCR cycle was carried out on the BioMark HD System (Fluidigm). The qPCR protocol consisted of a thermal mix (50 °C for 2 min, 70 °C for 30 min, 25 °C for 10 min) followed by UNG treatment and hot start (50 °C for 2 min, 95 °C for 10 min) then quantification (35 cycles of 96 °C for 15 sec then 60 °C for 1 min). Melting curve analysis was performed at the end of the qPCR protocol between 55 to 95 °C with 0.5 °C increments.

2.12 qPCR array data analysis

The Fluidigm Real-time PCR analysis software version 3.0.2 (Fluidigm) was used to check for amplification specificity using the melting curve data and to extract quantification cycle (Cq) data. PCR products with unacceptable melting curves were removed from further analysis. The Cq data was then imported into GenEx Pro version 5.4.2.128 (MultiD Analyses AB) for further analysis.

Genes that did not have valid data for more than 70% of the samples at each time point were excluded. Cq data of the remaining genes were pre-processed by first

correcting for primer efficiency using the following formula implemented in GenEx:

$$Cq_{E=100\%} = Cq \frac{\log(1 + E)}{\log(2)}$$

where E is the primer efficiency calculated by Real-time PCR miner (see section 2.11).

After correcting for primer efficiency, outliers among the technical triplicates were identified using Grubb's test (confidence level of 0.95 and cycle cutoff SD of 0.25) then removed from further analysis. The technical triplicates were averaged while those with only one data point remaining were flagged as missing data. The missing data were assigned the highest measured Cq for that gene plus one.

In the next step, processed Cq data were normalized to two reference genes (i.e. *eef1a111* and *rpl13a*):

$$\Delta Cq = Cq_{GOI} - \frac{1}{n} \sum_{i=1}^n Cq_{RGI}$$

where Cq_{GOI} is the efficiency corrected and processed Cq for the gene of interest. Cq_{RGI} is the average Cq (efficiency corrected and processed) of the two reference genes.

Finally, relative quantities (RQ: $2^{-\Delta Cq}$) of each gene per samples were imported into Partek Genomic Suite for further analysis using Principle Components Analysis (PCA), Hierarchical Clustering (HC) and Self-Organizing Map (SOM).

3 Results

3.1 Zebrafish has wide-ranging inter-family sex ratios

One of the classical methods used to investigate if a species utilizes the CSD system is to study progeny sex ratios from a large number of families. For CSD species, the proportion of males to females in a family is usually close to equal unless environmental factors or differential survival between the two sexes tilt the balance. For example, the Nile tilapia primarily uses a CSD system (XX/XY) that usually yields progeny sex ratios ranging between 40% to 60% males (Tessema *et al.* 2006).

To determine whether the primary SD mechanism of zebrafish is sex chromosome-based, we studied progeny sex ratio records of 62 families produced over a decade by various lab members and during this project. These families originated from different strains (i.e. AB, Tu and wild-type) and were all from pairwise crosses. The observed progeny sex ratio ranged from 4.8% to 97.3% males with the median at 51% (Figure 5). Three peaks were observed from the profile with the highest at 41-60% males and two smaller ones at 11-20% and 81-90% males. The additional two peaks in the male- (81-90%) and female-biased (11-20%) sex ratio range were not expected for CSD system. Therefore, the cumulative profile of progeny sex ratios suggested that zebrafish primary SD mechanism is not CSD-based.

Figure 5: Sex ratios of 62 zebrafish families showed a wide-ranging profile.

The observed family sex ratios showed a wide range from 4.8 to 97.3% males with the 41-50% and 51-60% male categories containing the most families.

3.2 Sex-linked markers identified by FluoMEP in guppy and rosy barb, but not in zebrafish

Further investigation on the presence or absence of highly differentiated sex chromosomes in zebrafish was performed by comparing male and female genomes using molecular approaches. The first method was to search for sex-linked DNA markers, an indication for highly differentiated sex chromosomes. For this purpose a ‘blind’ PCR-based screening technique called FluoMEP was developed in our lab (Chang *et al.* 2007). This new method combines the advantages of RAPD and

amplified fragment length polymorphism (AFLP). FluoMEP can identify large numbers of polymorphisms between genomes in a relatively short time.

Guppy (*Poecilia reticulata*) and rosy barb (*Puntius conchonius*), both with a male heterogametic (XX/XY) sex chromosomal system (Nanda *et al.* 1990; Kirankumar *et al.* 2003; Kirankumar and Pandian 2004), were used to test the ability of FluoMEP for detecting sex-specific polymorphic DNA markers in fish. For the guppy, 144 primer combinations were used for the FluoMEP screening. Using bulk segregant analysis, three male-specific markers were found (Figure 6) and all of them showed 100% sexing efficiency when tested on individual samples. The primer combinations yielding the markers were: common primer C117 with RAPD primer OPAL-16 (product size 395 bp), OPAO-08 (product size 490 bp) or OPAR-09 (product size 430 bp). For the rosy barb, 386 primer combinations were tested and two male-specific markers with 100% efficiency were identified (Figure 7). The first marker (product size 514 bp) was amplified by common primer C117 with RAPD primer OPL-20 and the second marker (product size 664 bp) was amplified by common primer C117 with Operon RAPD primer OPN-20. These results clearly indicated that the newly developed FluoMEP is capable of isolating sex-linked DNA markers from complex fish genomes with differentiated sex chromosomes. (For primer sequences of sex-linked markers please refer to Appendix 1).

After verifying that FluoMEP is suitable for isolating sex-linked markers from fish genomes, pooled gDNA samples from four adult male and female zebrafish siblings, respectively, were subjected to the same screening procedure. The pools

were screened with 258 primer combinations, but no sex-linked marker was identified. This suggests that there is no substantial structural difference between the male and female zebrafish genomes.

Figure 6: Male-specific FluoMEP markers from guppy.

Electrophoresis plots of FluoMEP result from bulked segregant analysis of pooled guppy (*Poecilia reticulata*) male and female gDNA samples. Three male-specific FluoMEP markers were isolated using common primer C117 with RAPD primer (A) OPAL-16, (B) OPAO-08 and (C) OPAR-09. Locations of the male-specific peak on the electrophoresis plot were marked by red box.

Figure 7: Male-specific FluoMEP markers from rosy barb.

Electrophoresis plots of FluoMEP result from bulked segregant analysis of pooled rosy barb (*Puntius conchonius*) male and female gDNA samples. Two male-specific markers were isolated with common primer C117 and RAPD primer (A) OPL-20 and (B) OPN-20. Locations of the male-specific peak on the electrophoresis plot were marked by red box.

3.3 The absence of a universal sex-linked CNVR supports lack of differentiated sex chromosomes in zebrafish

To further prove that no substantial differences exist between the two sexes of zebrafish, a more in-depth comparison of their genomes was required. Two custom-made array comparative genomic hybridization (aCGH) microarrays were designed for this purpose. Originally, the aCGH microarray was designed based on Zv7 (danRer5) of the zebrafish genome assembly, but during the course of this project the genome assembly was updated twice. Therefore, all aCGH microarray

probes were re-mapped onto the latest Zv9 (danRer7) genome assembly and results reported here were based on this latest genome assembly.

Our first aCGH microarray contained 108,417 probes on a Nimblegen 12 x 135K format chip (HX12). The probes were evenly distributed along the genome with median probe distance of 10,069 bp (see Appendix 2 for probes distribution on each chromosome) and a detection limit of 30 kb copy number variable region (CNVR). This study was performed on two families of the AB strain (i.e. AB1 and AB5) and two families of the Toh (wild-type) strain (i.e. Toh1 and Toh2). The sire's (paternal) genome was used as reference for each family; hence CNVRs were reported with copy numbers more or less than the sire. In total, 968 CNVRs (median length 40 kb) were identified from the four families (Figure 8). Among them, 257 CNVRs were found only in the AB strain and 297 CNVRs only in the Toh strain. None of the CNVRs identified showed sex-linked inheritance pattern in all four families tested.

Figure 8: Genomic map of the zebrafish copy number variation regions.

The genomic map showing regions found to have copy-number variation (CNV) using HX12 (~108K probes) aCGH microarray. CNVRs are represented by grey boxes on the chromosomes and size of the box is proportional to the CNVR length.

As the first custom-designed aCGH microarray was unable to identify CNVRs shorter than 30 kb, a second version based on Nimblegen 3 x 720K format with 661,018 probes (HX3) was made. Median probe distance of the second aCGH microarray was 1,688 bp with a detection limit of 5 kb CNVRs. An additional strain, called Tuebingen (Tu), was added to the previous two strains for this second aCGH study. Due to the higher cost of this microarray only one family from each of the AB, Toh and Tu strains was analyzed. Samples from each family consisted of the two parents, plus one male and one female offspring. The number of

CNVRs identified with this second version of aCGH microarray was more than a magnitude higher than that found with the first version (2,959 vs. 257). Even at the higher resolution no universal sex-linked CNVR was detected among the different strains. The result therefore indicated that there was no major difference between zebrafish male and female genomes, again suggesting the absence of a pair of well differentiated sex chromosomes in all the strains tested.

To validate the aCGH microarray data, gDNA-qPCR was performed using the sire as reference sample. A relative copy number of 0.5 or lower indicated that the copy number of the locus was less than that of the sire, while a relative copy number of two or more pointed to the opposite. Three CNVRs from the HX12 aCGH microarray results were randomly chosen for validation (Table 2). The gDNA-qPCR results for CNVR 1 and 2 (Figure 9A & B) were in agreement with the aCGH microarray results (Table 2). For CNVR3 the gDNA-qPCR result showed that only the dam had more copy number (Figure 9C), however the aCGH result showed that all the female samples had more copy number than the sire (Table 2). Overall, the gDNA-qPCR results validated two out of the three CNVRs indicating sufficient reliability of the aCGH microarray.

Table 2: Genomic location of the randomly chosen CNVRs for gDNA-qPCR validation and their HX12 aCGH microarray data for different families.

CNVR	CNVR			Family	aCGH microarray result				
	Chr	Start	End		M1	M2	Dam	Fe1	Fe2
1	8	30,299,989	30,300,095	Toh2	N	N	Less	Less	Less
2	7	2,938,738	2,938,894	Toh1	N	N	Less	Less	Less
3	1	57,569,066	57,569,185	AB5	N	N	More	More	More

Genomic location was based on Zv9 of the zebrafish genome assembly.

The aCGH results were reported using the sire result as reference.

N: neutral, similar copy number as sire.

M: male offspring and Fe: female offspring.

Figure 9: The relative copy number of three CNVRs chosen for gDNA-qPCR validation.

CNVRs were chosen from the HX12 aCGH microarray result (please refer to Table 2 for aCGH result). Three CNVRs from different families were chosen for the validation. Blue bars represent male samples and the red bars represent female samples. Sample 1 and 2 for both male and female were offspring used in the aCGH microarray experiment while sample 3 of both sexes were additional offspring used only for gDNA-qPCR validation. Relative copy number was calculated using the sire as reference. A relative copy number less than 0.5 indicated lower copy number compared to the sire, higher than 2 meant more copy number than the sire and between 0.5 to 1.5 meant similar copy number to the sire. (A) Validation of CNVR1 on chromosome 8 in family Toh2. (B) Validation of CNVR2 on chromosome 7 in family Toh1. (C) Validation of CNVR3 on chromosome 1 in family AB5.

3.4 Highly similar progeny sex ratios from repeated pairwise crossings indicate the presence of a GSD system in zebrafish

The breeding and molecular (i.e. FluoMEP and aCGH) data so far indicated that the primary SD mechanism of zebrafish is unlikely to be based on sex chromosomes. Therefore, the next experiment sought to investigate if zebrafish SD is genetically determined. If the primary SD mechanism was based on genetic factors, then the expected progeny sex ratios from the same parental pair bred repeatedly should be very similar. For this experiment, 18 pairs of zebrafish were chosen randomly for repeated (two times) pairwise breeding. In the first round of breeding, 10 pairs produced female-biased progeny sex ratios (0-37.5% males), while progeny sex ratios of five pairs were male-biased (68.3-88.1% males). The remaining three pairs produced near-equivalent progeny sex ratios (44.4-57.1% male). When the pairwise crosses were repeated, similar progeny sex ratios were observed between broods from the same breeding pair and was indicated by the high R^2 value of 0.89852 (Figure 10), despite having widely variable progeny sex ratios across the families. The largest progeny sex ratio difference observed was 15.4% for a female-biased family, which had 18.5% males in the first cross and 33.9% in the second. The average sex ratio difference observed among the 18 repeated crossing was 6.9%. The heritability of progeny sex ratios strongly suggested that genetic factor(s) play(s) an important role in zebrafish SD.

Figure 10: Repeated crossings of the same mating pair produced similar progeny sex ratios.

Eighteen pairs of zebrafish were crossed twice and the progeny sex ratios from the two different crosses were compared. The high R^2 value showed that in general sex ratios between broods from the same breeding pair were very similar. Based on the progeny sex ratio of the first cross, breeding pairs were categorized into female-biased group (red area), near-equivalent sex ratio group (yellow area) and male-biased group (blue area). Sex ratios (male %) for the 1st cross are reported on the X axis and for 2nd cross on the Y axis.

3.5 Response to selective full-sib pairwise breeding was in accordance with PSD system

Results from repeated pairwise breedings pointed toward potential influences of parental genotype on the progeny sex ratio. To further prove this observation, selective full-sib pairwise breeding was used to study the degree of parental genotypes influence on progeny sex ratio. Offspring from a family with female-biased sex ratio (4.8% males) were used to produce the H3 line in an attempt to fully deplete the males from this line by selective breeding (Figure 11A). Three male and three female full-sib zebrafish individuals from the F₁ generation were

randomly chosen for multifactorial pairwise crosses to produce six F₂ families (see Appendix 5 for sex ratio data of each family). The mean progeny sex ratio of the F₂ families was 23.9% males; well within the female-biased range (0-39% males). In the F₃ generation, the line was split into two by using offspring from two F₂ families with progeny sex ratio of 4.8% males (H3-21) and 28.6% males (H3-22) to produce two separate sub-lines (Figure 11A). Although the progeny sex ratios of the two chosen F₂ families differed close to 6-folds, their respective F₃ family-based mean progeny sex ratios showed no significant difference (*t-test*, $p = 0.19$). The mean progeny sex ratio of H3-21 line was 40.3% males and that of the H3-22 line was 32.6% males (Figure 11B). Although an all-female family was not obtained after three generations of selective breeding, the female-biased progeny sex ratio was maintained throughout.

The attempt was then changed to try to produce an all-male family and a male-biased line was created (D line). Five families were produced from pairwise breeding of ten randomly chosen siblings (five males and five females). The mean F₁ progeny sex ratio of the five families was 83.9% males (Figure 12A). Offspring from two families that had the highest proportion of males were then used to generate two sub-lines, namely D4 and D8 line. After four generations of selective breeding an all-male family was obtained and the other two F₄ families had extreme male-biased sex ratios (Figure 12B; see Appendix 5 for detailed sex ratio data).

To conduct further investigation into potential parental genotype influence, variation among progeny sex ratios was analyzed by calculating the coefficient of

variation (CV%) for families within the same generation. For the female-biased H3 line, it was observed that the F₂ generation had the highest variation at 87% (Figure 11C). This indicates a wide progeny sex ratio variation among the F₂ families. In contrast, the F₃ families produced by selected full-sib crossing had substantially lower progeny sex ratio variation at 18.7% and 38.8% for the H3-21 and H3-22 lines, respectively (Figure 11D).

Figure 11: Selective full-sib pairwise breeding of H3 line maintained female-biased sex ratio and decreased sex ratio variation.

(A) Progeny sex ratios of F₁ and F₂ generation. Each circle represents one family. Selected F₂ families used to generate F₃ were labeled red (H3-21) and green (H3-22). (B) F₃ brood sex ratios of the two sub-lines. H3-21 sub-line was produced by brooders from the F₂ family labeled by red color. H3-22 sub-line was produced by brooders from the F₂ family labeled by green color. (C) Progeny sex ratio variation of F₂ generation measured by coefficient of variation percentage (CV%). F₁ progeny sex ratio variation could not be calculated because only one family was produced. (D) F₃ progeny sex ratio variation of H3-21 and H3-22 lines.

Figure 12: Selective full-sib pairwise breeding of the male-biased D line.

(A) Progeny sex ratio of five random full-sib pairwise crosses to produce the male-biased D line. Two families with the highest proportion of males were selected to produce the F₂ generation. They were labeled by red (97.3%) and green (94.6%) circles. (B) Progeny sex ratios of D8 sub-line which was derived from the F₁ family that had 94.6% male (green circle). Selected F₂ and F₃ families used for subsequent selective breeding were indicated by the green circles. (C) Coefficient of variation (CV %) of D8 sub-line progeny sex ratios from F₁ to F₄ generations. (D) Progeny sex ratio of D4 sub-line derived from the F₁ family that produced 97.3% male (red circle). Selected F₂ family which offspring were used for mass cross was indicated by red circle. (E) Coefficient of variation (CV %) of D4 sub-line brood sex ratio from F₁ to F₄ generations. No progeny sex ratio variation CV% was calculated for F₂ generation as it was a mass cross.

As for progeny sex ratio variation in the male-biased D8 sub-line, the highest CV% was observed for the unselected F₁ generation at 22% (Figure 12C). Distinct reduction of progeny sex ratio variation to less than 10% was observed from F₂ to F₄ generations after selective breeding was performed (Figure 12C). To prove that the decreased progeny sex ratio variation was a result of selective breeding, a control mass cross (no selective breeding) was performed using the D4 sub-line. The F₁ generation produced by unselected brooders had progeny sex ratio variation of 22%, but after selective breeding F₂ progeny sex ratio variation was drastically decreased to 6.5% (Figure 12E). To produce the F₃ generation, a mass cross using brooders from a selected F₂ family (93.5% males) was carried out instead of pairwise breeding. The resulting F₃ brood had a progeny sex ratio of 84.8% males (Figure 12D). To produce the F₄ generation, pairwise breeding was again performed using randomly selected F₃ brooders (higher genotypes diversity as they were produced by mass crossing). The resulting progeny sex ratio variation for the F₄ generation increased about 4-folds to 25.4% (Figure 12E). Both F₁ and F₄ generations produced by unselected brooders (higher genotype diversity) had similar and higher progeny sex ratio variation. The distinct decrease of progeny sex ratio variation among families produced by full-sib brooders from selected sex biased family (lower genotype diversity) indicates a strong parental effect on zebrafish sex ratio.

Taken together, the ability to produce families with increased male- or female-biased sex ratios, the possibility to deplete females from one of the D line F₄ family and the reduction of progeny sex ratio variation by selective breeding indicated strong influence from parental genotypes.

3.6 A large number of transcripts showed dimorphic expression level during the gonadal differentiation period

The results so far showed that the zebrafish SD mechanism is genetically based with no sign of a CSD system. After resolving the general mechanism of zebrafish SD, the subsequent objectives of this project were to attempt to identify the genetic factors (genes) that may be involved in the process and to find out the approximate timing of SD. For identification of genetic factors, comparative gonadal transcriptome analysis was performed using a stable transgenic *Tg(vasa:vasa-EGFP)* zebrafish line (Krøvel and Olsen 2002). This line was previously reported to display sexually dimorphic vasa-EGFP fluorescence from as early as 28 dpf (Wang *et al.* 2007). Fish which had weak fluorescence developed as males, while the vast majority of those with very strong fluorescence became females (Wang *et al.* 2007). Therefore, gonadal samples were collected from individuals that displayed either strong or very weak fluorescence intensity starting from 28 dpf. Samplings continued at weekly intervals until 42 dpf and then an additional sample was taken upon reaching full maturity at 105 dpf (Figure 13). Gonadal transcriptomes were analyzed by a custom-designed expression microarray (see section 2.9) to search for genes with sexually dimorphic expression pattern throughout the whole period tested.

Figure 13: Sample collection scheme for expression analysis of gonad differentiation by microarray.

The graph shows fluorescence intensity of *Tg(vasa:vasa-EGFP)* transgenic zebrafish during the gonad differentiation period. Samples were collected at four time points (i.e. 28, 35, 42 and 105 dpf) which encompassed the majority of the differentiation period to mature ovary or testis. For each time point, samples were sorted according to their vasa-EGFP fluorescence intensity and only those that exhibited either very weak (future male) or strong (mostly future female) fluorescence were collected (denoted by the red circles). [Figure modified from (Wang *et al.* 2007)]

Transcriptome variation was analyzed by principle components analysis (PCA) with the microarray data and presented in a three-dimensional graphic representation (Figure 14). The 28 dpf samples clustered into two groups according to their fluorescence intensity, indicating that gonad differentiation has already started earlier. The separation between strong and weak fluorescence groups became progressively more distinct from 35 to 105 dpf. This is most likely due to the more differentiated state of the gonads. Among the various time points sampled, it was observed that zebrafish having the same fluorescence intensity at 35 and 42 dpf clustered closely together showing high similarity in their gonadal transcriptome profile.

Figure 14: PCA plot of the microarray data showed distinct divergence between the gonadal transcriptomes of strong (green) and weak (orange) vasa-EGFP fluorescence groups.

Samples analysed were collected from 28, 35, 42 and 105 dpf which overlapped a substantial period of the zebrafish gonad differentiation.

Similar to the PCA plot results, a progressive increase in the number of differentially expressed transcripts (DETs) between the strong and weak fluorescence groups (FDR adjusted p value ≤ 0.05 ; fold change ≥ 1.5) from 28 to 105 dpf was observed. At 28 dpf, there were 526 DETs between the strong and weak fluorescence groups. The numbers of DETs increased to 4,128 between the 35 dpf groups and to 6,051 between the 42 dpf groups. The highest number of DETs was observed between the 105 dpf groups with 12,037 transcripts. Altogether, there were 8,199 unique transcripts up-regulated in the strong

fluorescence groups and 8,325 unique transcripts up-regulated in the weak fluorescence groups. At most time points near equal numbers of transcripts were up-regulated between the strong and weak fluorescence groups, except at 28 dpf, whereby seven times more transcripts were up-regulated in the strong fluorescence group [460 (strong) vs. 66 (weak) transcripts].

In order to identify transcripts that showed consistent sexually dimorphic expression pattern from juvenile to adult stage, DETs up-regulated in either strong or weak fluorescence groups were compared across the four time points (Figure 15). Sixty nine transcripts were found to be consistently up-regulated in the strong fluorescence groups across all four time points (Figure 15A; for full list of the 69 transcripts please see Table 3). At the time of writing this thesis, 34 out of the 69 DETs had no known functions and among them six were from our gonadal EST library. Functional gene set enrichment analysis performed using the 69 DETs found significant enrichment for four GO processes (i.e. helicase activity, lyase activity, transition metal ion binding and zinc ion binding) and two KEGG pathways (progesterone-mediated oocyte maturation and cell cycle) (Table 4). On the other hand, no transcript showed consistent up-regulation in the weak fluorescence groups from 28 to 105 dpf (Figure 15B).

Table 3: List of the 69 DETs consistently up-regulated in strong (vs. weak) fluorescence groups from 28 to 105 dpf zebrafish.

Gene	Accession number	Gene description
<i>btg4</i>	NM_198121	B-cell translocation gene 4
<i>ca15b</i>	NM_213182	Carbonic anhydrase XV b
<i>ccna1</i>	NM_212818	Cyclin A1
<i>chia.4</i>	NM_200446	Chitinase, acidic.4
<i>csad</i>	NM_001007348	Cysteine sulfinic acid decarboxylase
<i>ctssb.1</i>	NM_001024409	Cathepsin Sb, tandem duplicate 1
<i>ddx41</i>	NM_201045	DEAD (Asp-Glu-Ala-Asp) box polypeptide 41
<i>grik1a</i>	XM_001924003	Glutamate receptor, ionotropic, kainate 1a
<i>gtf3ab</i>	NM_001089544	General transcription factor IIIA, b
<i>gtsf1</i>	ENSDARG00000000887	Gametocyte specific factor 1
<i>kpna7</i>	NM_213070	Karyopherin alpha 7 (importin alpha 8)
<i>loc100006609</i>	XM_001345268	Protein FAM55C-like
<i>loc100006634</i>	XM_001345287	Protein FAM55C-like
<i>loc100150238</i>	XM_001923225	Lysosomal thioesterase PPT2-like
<i>loc100331651</i>	XM_002667287	Zona pellucida sperm-binding protein 4-like
<i>loc100332229</i>	XM_003197608	Histone H2A type 1-A-like
<i>loc100333071</i>	XM_002667105	Matrix metalloproteinase-16-like
<i>loc100334979</i>	XM_002666960	Tudor domain-containing protein 12-like
<i>loc100536102</i>	XM_003200303	Tudor domain-containing protein 6-like
<i>loc556271</i>	XM_678169	Protein PAT1 homolog 2-like
<i>loc566400</i>	XM_689667	Protein FAM55C-like
<i>mcm3l</i>	NM_001171068	MCM3 minichromosome maintenance deficient 3 (S. cerevisiae), like
<i>mos</i>	NM_205580	v-mos Moloney murine sarcoma viral oncogene homolog
<i>nanog</i>	NM_001098392	Nanog homeobox
<i>pdcd7</i>	NM_001003415	Programmed cell death 7
<i>prmt6</i>	NM_001163988	Protein arginine methyltransferase 6
<i>rbpms2a</i>	NM_001002409	RNA binding protein with multiple splicing 2a
<i>retsatl</i>	NM_001040004	Retinol saturase (all-trans-retinol 13,14-reductase) like
<i>snopc1a</i>	NM_200263	Small nuclear RNA activating complex, polypeptide 1a
<i>st6galnac1.2</i>	NM_001082934	ST6 (alpha-N-acetyl-neuraminyl-2,3-beta-galactosyl-1,3)-N-acetylglactosaminide alpha-2,6-sialyltransferase 1, tandem duplicate 2
<i>tph1a</i>	NM_178306	Tryptophan hydroxylase 1 (tryptophan 5-monooxygenase) A
<i>trim35-29</i>	NM_001007347	Tripartite motif containing 35-29
<i>wnt8-2</i>	NM_001020801	Protein Wnt-8a ORF2
<i>zc3h7a</i>	NM_213434	Zinc finger CCCH-type containing 7A
<i>zfand6</i>	NM_205760	Zinc finger, AN1-type domain 6
<i>CO350393</i>	CO350393	EST
<i>CO351530</i>	CO351530	EST
<i>EV559355</i>	EV559355	EST
<i>EV603839</i>	EV603839	EST
<i>EV604784</i>	EV604784	EST
<i>EX156489</i>	EX156489	EST
<i>cu929143.1</i>	ENSDARG000000087392	Uncharacterized
<i>loc100001286</i>	XM_001337464	Uncharacterized
<i>loc100004563</i>	XM_001337068	Uncharacterized
<i>loc100332459</i>	XM_002662036	Uncharacterized
<i>loc100332474</i>	XM_002660821	Uncharacterized
<i>loc100534739</i>	XM_003199490	Uncharacterized
<i>loc100535616</i>	XM_003201245	Uncharacterized
<i>loc100536086</i>	XM_003200568	Uncharacterized
<i>loc100536087</i>	XM_003200607	Uncharacterized
<i>loc100536640</i>	XM_003200581	Uncharacterized
<i>loc556026</i>	XM_678717	Uncharacterized
<i>si:ch211-119o8.6</i>	NM_001114740	Uncharacterized
<i>si:ch211-188c16.1</i>	NM_001177447	Uncharacterized
<i>si:ch211-225p5.3</i>	NM_001162847	Uncharacterized
<i>si:ch211-262i10.2</i>	ENSDARG000000092538	Uncharacterized
<i>si:dkey-22a1.3</i>	NM_001017650	Uncharacterized
<i>si:dkey-88l16.3</i>	XM_001337253	Uncharacterized
<i>wu:fi09b08</i>	XM_002663529	Uncharacterized
<i>wu:fi34b01</i>	XM_001338062	Uncharacterized
<i>wu:fi42e03</i>	XM_001340211	Uncharacterized
<i>zgc:110816</i>	NM_001014331	Uncharacterized
<i>zgc:152652</i>	NM_001045340	Uncharacterized
<i>zgc:171474</i>	NM_001114710	Uncharacterized
<i>zgc:171977</i>	NM_001102647	Uncharacterized
<i>zgc:173556</i>	NM_001109742	Uncharacterized
<i>zgc:173837</i>	NM_001111233	Uncharacterized
<i>zgc:66432</i>	NM_199838	Uncharacterized
<i>zgc:66449</i>	NM_199801	Uncharacterized

Figure 15: Comparison of up-regulated transcripts in strong or weak fluorescence groups between 28 to 105 dpf.

DETs were first separated into either up-regulated in strong or weak fluorescence group then compared using venn diagram to identify those with consistent expression pattern across all time points. **(A)** Sixty nine transcripts were up-regulated in the strong fluorescence groups from 28 to 105 dpf. **(B)** No transcript was found to have common expression pattern in the weak fluorescence groups across all time points. Numbers of DETs are indicated in brackets.

Table 4: GO terms and KEGG pathways enriched in the list of 69 DETs up-regulated in strong fluorescence groups from 28 to 105 dpf zebrafish.

Description	ID	Enrichment score	Enrichment <i>p</i> value
<u>GO term</u>			
Helicase activity	GO:0004386	3.746	0.024
Lyase activity	GO:0016829	3.593	0.028
Transition metal ion binding	GO:0046914	3.592	0.028
Zinc ion binding	GO:0008270	3.239	0.039
<u>KEGG pathway</u>			
Cell cycle	dre04110	4.068	0.017
Progesterone-mediated oocyte maturation	dre04914	4.497	0.011

To validate the microarray data, qPCR was performed on four randomly chosen transcripts (i.e. *lin28a*, *qik2*, *wu:fd16d01* and *zgc:162969*) from the 105 dpf DET list. When the qPCR results were compared with the microarray results, all of them showed an identical trend of expression across all four time points (Figure 16). However, for *lin28a* at 35 dpf, *qik2* at 105 dpf, *wu:fd16d01* at 35 dpf and *zgc:162969* at 35 dpf, either the qPCR or microarray result did not show statistically significant difference. Other than these minor disagreements, the majority of the microarray data were validated by the qPCR results.

Figure 16: Validation of expression microarray result by qPCR.

Expression levels of (A) *lin28a*, (B) *qik2*, (C) *wu:fd16d01* and (D) *zgc:162969* were measured by qPCR (blue bar) and compared to the microarray result (yellow bar). Fold change was calculated by comparing weak to strong vasa-EGFP fluorescence groups. Statistical analysis was calculated using ANOVA for microarray data and Student's *t*-test for qPCR data, significant difference ($p \leq 0.05$) was indicated by an asterisk. Transcript with no significant difference for both methods was labeled as NS.

3.7 Enrichment of functional gene sets from GO and KEGG databases

In addition to searching for individual transcripts of interest, the microarray data were also used for identifying functional gene sets with differential expression level between the strong and weak fluorescence groups. For this purpose the more powerful ANOVA based method implemented in Partek GS was used (see section 2.10). The number of gene sets in the GO database that showed significant differences (FDR adjusted $p \leq 0.1$ and average fold change ≥ 1.2) between strong and weak fluorescence groups were 74 at 28 dpf, 1,406 at 35 dpf, 1,838 at 42 dpf and 2,127 at 105 dpf. All time points, except 28 dpf, had gene sets with significant difference from the three GO domains (i.e. biological process, cellular component and molecular function) of the GO database (Figure 17). The weak fluorescence group at 28 dpf did not have any gene set from the GO molecular function domain that showed significant up-regulation compared to the strong fluorescence group. Seven GO biological processes and seven GO molecular function gene sets were both found to be consistently up-regulated from 28 to 105 dpf (Table 5). No common gene set with consistent expression pattern was found among the four time points in the weak fluorescence group.

Figure 17: Comparison of GO terms found to be significantly up-regulated in the strong or weak fluorescence groups from all time points.

The three domains covered by GO database (i.e. biological process, cellular component and molecular function) were represented in the rows. Left panels are venn diagrams of GO terms up-regulated in the strong fluorescence groups while the right panels are from the weak fluorescence groups. No molecular function term was identified for the 28 dpf weak fluorescence group. Numbers of GO terms with significant differential expression are indicated in the bracket.

Table 5: GO terms that were up-regulated in the strong fluorescence group from 28 to 105 dpf.

GO ID	GO domain	GO Description
GO:0000079	biological process	Regulation of cyclin-dependent protein serine/threonine kinase activity
GO:0001541	biological process	Ovarian follicle development
GO:0006865	biological process	Amino acid transport
GO:0015849	biological process	Organic acid transport
GO:0016458	biological process	Gene silencing
GO:0031047	biological process	Gene silencing by RNA
GO:0046942	biological process	Carboxylic acid transport
GO:0000339	molecular function	RNA cap binding
GO:0000340	molecular function	RNA 7-methylguanosine cap binding
GO:0003678	molecular function	DNA helicase activity
GO:0003729	molecular function	mRNA binding
GO:0008026	molecular function	ATP-dependent helicase activity
GO:0008474	molecular function	Palmitoyl-(protein) hydrolase activity
GO:0070035	molecular function	Purine NTP-dependent helicase activity

In addition to the GO database, functional gene sets from KEGG pathways database were also used. No KEGG pathway exhibited significant difference between the weak and strong fluorescence groups at 28 dpf. Twelve pathways were up-regulated in the strong fluorescence groups from 35 to 105 dpf, while only five pathways showed consistent up-regulation in the weak fluorescence group across all the same time points (Figure 18; Table 6).

Figure 18: Comparison of KEGG pathways up-regulated in strong or weak fluorescence groups.

Twelve pathways were up-regulated in the strong fluorescence groups from 35 to 105 dpf while only five pathways were up-regulated in the weak fluorescence groups across the same time points.

Table 6: KEGG pathways with consistent expression pattern from 35 to 105 dpf zebrafish.

KEGG ID	KEGG Description	Up-regulated in fluorescence group
dre03060	Protein export	Strong
dre03430	Mismatch repair	Strong
dre03440	Homologous recombination	Strong
dre03450	Non-homologous end-joining	Strong
dre03460	Fanconi anemia pathway	Strong
dre03008	Ribosome biogenesis in eukaryotes	Strong
dre04120	Ubiquitin mediated proteolysis	Strong
dre04122	Sulfur relay system	Strong
dre04130	SNARE interactions in vesicular transport	Strong
dre04141	Protein processing in endoplasmic reticulum	Strong
dre03013	RNA transport	Strong
dre03022	Basal transcription factors	Strong
dre04060	Cytokine-cytokine receptor interaction	Weak
dre04260	Cardiac muscle contraction	Weak
dre04512	ECM-receptor interaction	Weak
dre04672	Intestinal immune network for IgA production	Weak
dre04744	Phototransduction	Weak

3.8 Expression profile of a selected set of genes revealed transcriptomic differences in zebrafish gonads from 28 to 42 dpf

Using 11 of the 69 DETs showing consistent expression pattern from 28 to 105 dpf in the microarray experiment and nine sex candidate genes, a qPCR array was designed for gene expression profiling (Table 7; for primers details see Appendix 4). This tool was later used to determine and compare the gonadal transcriptome of ‘unsorted’ samples from earlier stages (7, 14 and 21 dpf) in order to find out when the zebrafish gonad starts to develop into testis or ovary.

The ability of these 20 genes to sex zebrafish was first tested on 28, 35 and 42 dpf *Tg(vasa:vasa-EGFP)* transgenic zebrafish, in which fluorescence intensity provided indication of the future sex. The vasa-EGFP fluorescence intensity of each sample was checked visually under stereo-microscope and recorded. Due to exclusion of genes, which did not have valid data in at least 70% of the samples, only 17 genes were analysed for 28 dpf. Samples were separated into various groups by gene expression profiling using HC, SOM and PCA (see Table 8 for mean relative quantity of the twenty genes).

Table 7: Genes and transcripts selected for qPCR array from the microarray result and published studies.

Gene Symbol	Accession Number	Description
<u>Differentially expressed transcripts</u>		
<i>CO351530</i>	CO351530	Novel (EST)
<i>EV559355</i>	EV559355	Novel (EST)
<i>csad</i>	NM_001007348	cysteine sulfinic acid decarboxylase
<i>gtf3ab</i>	NM_001089544	general transcription factor IIIA, b
<i>mcm3l</i>	NM_001171068	MCM3 minichromosome maintenance deficient 3 (<i>S. cerevisiae</i>), like
<i>retsatl</i>	NM_001040004	retinol saturase (all-trans-retinol 13,14-reductase) like
<i>si:ch211-119o8.6</i>	NM_001114740	Novel
<i>snopc1a</i>	NM_200263	small nuclear RNA activating complex, polypeptide 1a
<i>zgc:66449</i>	NM_199801	Novel
<i>loc100334979</i>	XM_002666960	Novel (predicted from genome sequences)
<i>wu:fi34b01</i>	XM_001338062	Novel (predicted from genome sequences)
<u>Sex candidate genes</u>		
<i>amh</i>	NM_001007779	anti-Mullerian hormone
<i>cyp19a1a</i>	NM_131154	cytochrome P450, family 19, subfamily A, polypeptide 1a
<i>dmrt1</i>	NM_205628	doublesex and mab-3 related transcription factor 1
<i>dnd</i>	NM_212795	dead end
<i>foxl2</i>	NM_001045252	forkhead box L2
<i>ift57</i>	NM_001001832	intraflagellar transport protein 57
<i>pou5f1</i>	NM_131112	POU domain, class 5, transcription factor 1
<i>vtg5</i>	NM_001025189	vitellogenin 5
<i>zp2</i>	NM_131330	zona pellucida glycoprotein 2

Table 8: Mean relative quantity of the 20 genes measured by qPCR array for 28, 35 and 42 dpf zebrafish juveniles.

Gene	42 dpf		35 dpf			28 dpf		
	Group 1	Group 2	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3
<i>amh</i>	7.413E-04	1.258E-01	1.416E-03	1.203E-01	4.703E-03	1.483E-04	5.877E-05	1.017E-04
<i>csad</i>	7.693E-02	2.543E-02	4.692E-02	1.285E-02	2.084E-02	1.561E-02	1.400E-02	1.469E-02
<i>cyp19a1a</i>	3.374E-03	8.109E-04	8.319E-03	7.921E-04	3.035E-03	4.966E-04	3.379E-04	1.417E-04
<i>dmrt1</i>	1.142E-02	7.734E-02	9.729E-03	6.006E-02	1.596E-02	5.802E-04	1.494E-04	5.119E-04
<i>dnd</i>	1.166E-01	8.821E-02	1.236E-01	1.376E-01	1.010E-01	3.058E-03	7.112E-05	8.171E-04
<i>ev559355</i>	1.294E-01	6.279E-03	9.627E-02	1.077E-02	3.250E-03	N.D.	N.D.	N.D.
<i>loc100334979</i>	1.048E-01	7.032E-02	8.764E-02	7.750E-02	1.053E-01	3.725E-03	2.345E-04	1.426E-03
<i>foxl2</i>	2.748E-03	9.256E-04	5.526E-03	2.507E-03	8.183E-03	2.314E-04	4.434E-05	1.625E-04
<i>gtf3ab</i>	8.181E-01	4.842E-02	4.751E-01	7.180E-02	3.059E-02	9.822E-03	7.092E-05	1.538E-04
<i>ift57</i>	1.643E-02	8.890E-03	1.269E-02	1.345E-02	1.108E-02	2.547E-03	2.136E-03	2.606E-03
<i>mcm3l</i>	1.349E-01	8.128E-03	1.062E-01	1.284E-02	2.904E-03	9.076E-04	5.813E-05	6.957E-05
<i>pou5f1</i>	1.626E-01	7.203E-03	8.220E-02	9.781E-03	3.154E-03	2.798E-04	3.828E-05	3.923E-05
<i>restat1</i>	1.567E+00	7.362E-02	1.174E+00	7.517E-02	2.627E-02	2.402E-02	1.128E-04	3.410E-04
<i>si:ch211-119o8.6</i>	1.046E-01	1.977E-02	7.941E-02	2.536E-02	1.352E-02	2.406E-03	1.491E-04	1.248E-04
<i>snapc1a</i>	5.745E-02	2.063E-03	5.330E-02	2.563E-03	1.995E-03	2.684E-03	1.458E-04	1.229E-04
<i>vtg5</i>	3.711E-05	4.988E-05	3.097E-05	8.455E-05	3.825E-05	N.D.	N.D.	N.D.
<i>co351530</i>	4.231E-01	4.325E-02	2.631E-01	4.509E-02	1.489E-02	2.291E-03	8.310E-05	1.229E-04
<i>wu:fi34b01</i>	3.329E-01	2.899E-02	2.013E-01	3.623E-02	1.246E-02	N.D.	N.D.	N.D.
<i>zgc:66449</i>	9.514E-02	7.634E-03	7.054E-02	1.110E-02	8.766E-03	1.660E-03	1.472E-05	2.189E-05
<i>zp2</i>	9.375E+00	3.275E-01	7.400E+00	3.428E-01	6.626E-02	1.293E-01	7.315E-05	3.477E-04

N.D.: No valid data

Figure 19: Expression profiling of zebrafish genes associated with gonad development revealed an intermediate group at 28 and 35 dpf.

Left panels (A, C & E): Results of HC presented in dendrogram including fluorescence intensity and SOM grouping of each sample. Samples' vasa-EGFP fluorescence intensity checked by visual sorting were indicated next to the dendrogram (strong fluorescence: green; weak fluorescence: light blue and intermediate fluorescence: orange). SOM groupings were indicated in the open box with group number. The last column indicates sample ID. **Right panels (B, D & F):** 3D plot integrating results from PCA and SOM. Fluorescence intensity of each sample was labeled by either green for strong fluorescence, light blue for weak fluorescence or orange for intermediate fluorescence. SOM groupings were represented by the eclipse surrounding the samples with grouping number beside it.

Fourteen juvenile zebrafish individuals at 42 dpf age were selected for this experiment with six samples having weak (future males) and eight with strong (future females) vasa-EGFP fluorescence. All twenty genes yielded valid data for every sample. The three different methods used for gene expression profiling grouped the samples into two distinct groups (six samples in one group and the rest in another group) with identical results and corresponded well to the samples' fluorescence intensity (Figure 19A & B). Expression pattern for most sex candidate genes were as expected, except for *vtg5* which showed higher expression in the weak fluorescence group (Figure 19A).

For the 35 dpf time point, four samples with weak fluorescence and ten samples showing strong fluorescence were analyzed. All twenty genes produced valid data. HC analysis of the 35 dpf samples clustered them into two major clades which corresponded to the samples' vasa-EGFP fluorescence intensity recorded visually. The molecular method which was more sensitive further sub-divided the weak fluorescence samples into two minor clades (Figure 19C). The HC groupings were confirmed by SOM and PCA plot, which also separated the weak fluorescence samples in two distinct clusters (Figure 19C & D). In total, three groups were identified within the 35 dpf samples using the three molecular methods. A closer examination of the expression level of two “pro-male” genes (*amh* and *dmrt1*) as well as two “pro-female” genes (*pou5f1* and *zp2*) revealed the status of Group 2 and 3 samples. For “pro-male” genes, both Group 3 samples (weak fluorescence) and Group 1 (strong fluorescence) samples had similar mean expression levels, which were lower than Group 2 (weak fluorescence) samples (Figure 20A & B). On the other hand, mean expression levels of “pro-female” genes for both weak

fluorescence groups (Group 2 and 3) were lower than the strong fluorescence group (Group 1) samples (Figure 20C & D). In general, mean expression levels of “pro-male” genes in Group 3 (weak fluorescence) samples were similar to the strong fluorescence (Group 1), whereas the “pro-female” ones were closer to those of the weak fluorescence (Group 2) samples. This suggests that Group 3 samples are likely to be future males that were transforming from juvenile ovary to testis.

Figure 20: Expression levels of “pro-male” and “pro-female” genes at 35 dpf confirmed the presence of an intermediate fluorescence group.

The relative quantity ($2^{-\Delta Cq}$) of “pro-male” genes (A) *amh* and (B) *dmrt1* were lower in both Group 1 and 3 compared to Group 2. For the “pro-female” genes (C) *pou5f1* and (D) *zp2* expression level for Group 2 and 3 were lower than Group 1. Overall, Group 3 samples had lower expression for all the four genes tested. Samples with weak fluorescence were labeled in blue while strong fluorescence samples were labeled in green. Horizontal lines represent mean relative quantity for each group. Samples were grouped by expression profiling of a selected set of genes.

For the 28 dpf samples, three genes (i.e. *wu:fi34b01*, *ev559355* and *vtg5*) did not produce valid data for more than 70% of the samples, most probably due to their low expression level at that stage. Hence, these genes were not used for the transcriptome-based grouping of the samples. Sorting the 28 dpf zebrafish according to their vasa-EGFP fluorescence signal by visual analysis resulted in three groups (i.e. strong, weak and intermediate). Using HC, SOM and PCA, the samples were also clustered into three groups. In general, the composition of groups produced by molecular and fluorescence-based sorting was very similar (Figure 19E & F). However, the intermediate fluorescence group was only revealed by the more sensitive molecular method. All the strong fluorescence samples were sorted into one group and the weak fluorescence samples into two groups similar to the 35 dpf samples. Out of the three intermediate fluorescence samples identified visually, two of them clustered with either of the weak fluorescence groups, while the remaining one clustered with the strong ones (Figure 19E & F). Expression level of two “pro-male” (*amh* and *dmrt1*) and two “pro-female” (*pou5f1* and *cyp19a1a*) genes were also examined for the 28 dpf samples. Group 1 samples showed higher mean expression levels for all four selected genes when compared to Group 2 or 3 samples (Figure 21). Between the two weak fluorescence groups, Group 3 had higher (but not statistically significant) mean expression level than Group 2 for both “pro-male” genes *amh* (Figure 21A) and *dmrt1* (Figure 21B). For the “pro-female” genes, Group 3 had similar mean expression level for *pou5f1* (Figure 21C) but lower mean expression level for *cyp19a1a* (Figure 21D; Student’s *t-test* $p \leq 0.05$) when compared to Group 2. This suggests that Group 2 samples were most probably undergoing gonad transformation.

Taken together, the result showed that it is possible to sex juvenile zebrafish from 28 to 42 dpf using the expression profile of a carefully selected set of genes from microarray data obtained in this project or published data. Samples showing strong or weak fluorescence intensity at the three time points were correctly identified and grouped together. Furthermore, molecular-based methods were more sensitive and could sex 28 dpf zebrafish with indiscernible vasa-EGFP fluorescence intensity such as samples with intermediate vasa-EGFP fluorescence intensity.

Figure 21: Expression profiles of two “pro-male” and “pro-female” genes at 28 dpf.

Relative quantity (2^{-ΔCq}) of “pro-male” genes (A) *amh* and (B) *dmrt1* as well as “pro-female” genes (C) *pou5f1* and (D) *cyp19a1a* were measured by qPCR. Samples were grouped according to the results of gene expression profiling. Samples with weak fluorescence were labeled in blue, strong fluorescence samples in green and intermediate fluorescence in orange. Horizontal lines represent mean relative quantity for each group.

3.9 The expression pattern of a selected gonadal gene set indicated that sex determination in zebrafish started prior to 7 dpf

After verifying the ability to sex juvenile zebrafish using expression profiling of a selected set of genes, the method was used on zebrafish larvae in an attempt to find out when the gonad differentiation started. Using the same family as in juvenile sexing, trunk samples from zebrafish larvae were collected randomly at 7, 14 and 21 dpf. Although *Tg(vasa:vasa-EGFP)* transgenic zebrafish was used, no prior sorting was performed due to technical difficulties of quantifying fluorescence intensity visually on small fish. Compared to the juveniles (28 to 42 dpf) analyzed in the previous round, the number of genes that produced valid qPCR data for more than 70% of the larval samples (7 to 21 dpf) were lesser. At 21 dpf, valid data were obtained from nine genes, while 14 dpf samples yielded eight genes and 7 dpf seven genes. Nevertheless, the three molecular methods (i.e. HC, SOM and PCA) were able to sort all zebrafish larvae at the various time points into different groups. The larval samples were grouped according to HC and PCA results because these two methods showed a good agreement. Using expression profiling, the 21 and 14 dpf zebrafish larvae were clustered into three different groups (Figure 22A-D), while 7 dpf larvae were separated into two groups (Figure 22E & F). The result showed that zebrafish gonadal transcriptome started to diverge as early as 7 dpf. However, differences in gene expression levels between the groups at larval stages were mostly not statistically significant (see Table 9 for mean relative quantity for the genes at 7 to 21 dpf).

Figure 22: Expression profiling of a selected set of genes during larval stages in zebrafish showed the divergence of gonadal transcript levels starting from 7 dpf.

Left panels (A, C & E): Larval samples were grouped by HC into various clades using subset of the selected genes in qPCR array. The SOM groupings were also indicated beside the dendrogram in the open box with group number. **Right panels (B, D & F):** 3D plots of PCA and SOM results. The SOM grouping results were labeled by different colors. In general, it was observed that the HC groupings were in good agreement with the PCA results. Final grouping of the samples are indicated by the numbers beside each cluster of samples.

Table 9: Mean relative quantity of the 20 genes measured by qPCR array for 7, 14 and 21 dpf zebrafish larvae.

Gene	21 dpf			14 dpf			7 dpf	
	Group 1	Group 2	Group 3	Group 1	Group 2	Group 3	Group 1	Group 2
<i>amh</i>	8.167E-05	1.512E-04	1.880E-04	N.D.	N.D.	N.D.	N.D.	N.D.
<i>csad</i>	1.040E-02	7.001E-03	1.825E-02	6.264E-03	8.009E-03	5.430E-03	5.839E-03	5.049E-03
<i>cyp19a1a</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>dmrt1</i>	4.866E-05	3.877E-04	2.382E-04	3.021E-04	2.880E-04	1.205E-04	1.780E-04	4.056E-04
<i>dnd</i>	5.665E-05	2.847E-04	2.074E-04	3.520E-04	1.551E-04	6.007E-05	2.243E-04	1.028E-04
<i>ev559355</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>loc100334979</i>	5.025E-05	3.118E-04	3.808E-04	6.480E-04	1.215E-04	2.344E-04	2.995E-04	2.376E-04
<i>foxl2</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>gtf3ab</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>ift57</i>	2.480E-03	1.676E-03	2.978E-03	1.801E-03	2.217E-03	1.788E-03	1.309E-03	2.036E-03
<i>mcm3l</i>	7.105E-05	5.743E-05	1.240E-05	3.806E-05	3.527E-05	5.413E-05	1.847E-04	1.535E-04
<i>pou5f1</i>	6.260E-05	5.873E-05	2.170E-05	5.500E-05	2.394E-05	4.929E-05	3.645E-05	1.002E-04
<i>restat1</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>si:ch211-119o8.6</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>snpc1a</i>	1.458E-04	6.037E-05	6.890E-05	1.534E-04	3.224E-05	4.279E-05	N.D.	N.D.
<i>vtg5</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>co351530</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>wu:fi34b01</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>zgc:66449</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.
<i>zp2</i>	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.	N.D.

N.D.: No valid data

4 Discussion

4.1 Zebrafish sex is determined by genetic factor(s)

Despite being a model organism for developmental biology study for a few decades, the SD mechanism of zebrafish is still elusive. The first task of this project was to find out which of the two broad categories of SD mechanism (namely GSD or ESD) does zebrafish belongs to.

Several experiments were carried out in this project to find out if genetic factors are governing sexual development in zebrafish. One of the experiments was selective breeding using fish from families with skewed sex ratios (see section 3.5). Full depletion of females was obtained in one F₄ family (male-biased D line), but attempts to produce an all-female family were not successful. The exact reason for this is unknown and could be due to the lack of strong “pro-female” factors in the population of zebrafish used. It was also observed that selective breeding was able to maintain biased sex ratio for both sexes and reduced sex ratio variation among families within the same generation. Together these two observations revealed that sex is a heritable trait in zebrafish, which is a strong indication for a GSD system.

The next study compared progeny sex ratios of broods from repeated pairwise breeding to determine the influence of environmental factors on zebrafish sex ratio. Several pairs were bred twice on different occasions and environmental conditions were deliberately not kept similar. Variations in the offspring rearing

density, amount of food given (which directly affect growth rate) and ambient temperature were expected. Nonetheless, the two broods produced by the same breeding pair from either the sex-biased or equivalent sex ratio families had very similar progeny sex ratios (see section 3.4). This is another clear indication that sex is a heritable trait and also argued against zebrafish as an ESD species. Although zebrafish reared at high temperature (35 to 37 °C) during the gonad transformation period (18 to 28 dpf) resulted in male-biased sex ratio (see section 1.5), exposing 1 to 5 dpf larvae to 35 °C (Abozaid *et al.* 2011) or 37 °C (data not shown) resulted in high mortality and low hatching rate. The fact that zebrafish larvae cannot survive at elevated water temperature provided additional indication that zebrafish is not an ESD species with temperature-dependent SD mechanism. Furthermore, the typical temperature range (26 to 32 °C) of zebrafish natural habitat (Engeszer *et al.* 2007) does not induce masculinization. Therefore, we support the conclusion of Piferrer's group that zebrafish does not have temperature-dependent sex determination system, but is a GSD species with temperature effect (Ospina-Alvarez and Piferrer 2008).

4.2 Zebrafish is not a CSD species

Having established that zebrafish sex was determined by genetic factor(s), the next task was to find out the type of GSD system used. To investigate if CSD system is present in the zebrafish, two different molecular platforms were developed during this project to look for substantial structural differences between zebrafish male and female genomes. The first molecular method used was a new PCR-based genomic screening assay called FluoMEP that was developed in collaboration with other lab members (Chang *et al.* 2007). Analysing the zebrafish with FluoMEP did

not yield any universal sex-linked marker. On the other hand, male-specific FluoMEP markers were obtained from two fish species (i.e. guppy and rosy barb) that utilized CSD system (see section 3.2). This indicates that no major structural chromosomal difference exists between the two zebrafish sexes. However, there are some inadequacies in the FluoMEP assay such as it allows only partial control of the targeted sites for screening by using different common primer. In addition the targeted common motif might not be evenly distributed in the genome.

To conclusively rule out substantial structural difference uniformly present between all male and female zebrafish genomes, aCGH which allows comparison of the whole genome at high resolution was used. For this project, two types of aCGH microarrays were custom designed that can detect CNVRs of 30 kb and 5 kb, respectively. This size range should be sufficient for the objective of this experiment as the length of sex determining region surrounding the primary switch is expected to be more than 30 kb, for example the sex determining region showing suppressed recombination on the medaka Y chromosome is around 250 kb (Matsuda *et al.* 2002). A total of five zebrafish families from three strains were screened, but none of them yielded any universal sex-linked CNVR (see section 0). A potential caveat for the aCGH approach is that the sex-linked CNVR could be located in the ca. 55.4 Mb of sequences (ca. 3.9% of the total zebrafish genome sequences obtained) that cannot be mapped onto any chromosome in the current zebrafish genome assembly (Zv9; Genbank Assembly ID GCA_000002035.2). However, results obtained from aCGH and FluoMEP in this project as well as the ten karyotyping studies, which did not find any cytogenetically recognizable sex chromosomes (Post 1964; Endo and Ingalls 1968; Rishi 1976; Schreeb *et al.* 1993;

Pijnacker and Ferwerda 1995; Daga *et al.* 1996; Gornung *et al.* 1997; Amores and Postlethwait 1998; Ueda and Naoi 1999; Phillips and Reed 2000), provide an indication that no substantial, universal difference exists between the two sexes' genomes in all zebrafish strains.

The absence of a heteromorphic chromosomal pair and the lack of substantial genomic differences between the two sexes both strongly suggest that zebrafish does not have highly differentiated sex chromosomes. However, the lack of structural differences between the two sexes cannot rule out the possibility that CSD system is used by zebrafish. In fact, size dimorphic sex chromosomes are very rare in teleosts and some fish species utilizing CSD system do not have cytogenetically recognizable sex chromosomes (see section 1.4.1). To overcome this shortcoming, studying progeny sex ratios from a large number of different pairwise crosses can be used to find out if the CSD system is being used (Devlin and Nagahama 2002). Species whose primary SD mechanism is CSD (e.g. Nile tilapia) will only have families with near equivalent sex ratio (Figure 23A). A substantial portion (nearly 40%) of the 62 zebrafish families tested exhibited male- or female-biased sex ratio, which clearly showed that zebrafish is not a CSD species. Furthermore the progeny sex ratios observed in this project showed a profile (Figure 5) that was very similar to species that utilizes sex determination based on multiple chromosomes such as Florida red tilapia (Figure 23B). Our observation is supported by other reports, which also showed that zebrafish sex is incompatible with a simple CSD system by analyzing progeny sex ratios from either whole genome manipulation or breeding of sex-reversed fish (see sections 1.4.2 and 1.4.3). However, one study suggested that zebrafish had a 'female

dominance chromosomal based sex determination mechanism' (Tong *et al.* 2010) which is identical to the female heterogametic system (ZW/ZZ). They identified "super females" among the F₁ fish from crossing neo-male zebrafish (obtained by hormonal sex reversal) to normal female. Using the "super females" they managed to generate families that were severely female-biased that suggests the existence of a female heterogametic system. Unfortunately, there is no indication in the publication whether the authors have attempted to identify the proposed sex chromosomes (or any of their markers). The sex ratios reported also do not conform to the female heterogametic system (see section 1.4.3 and Figure 1 of (Tong *et al.* 2010). The F₁ sex ratios of 61.2-61.7% female after hormone treatment were below the expected 75% of females. Furthermore it was expected that one third of the F₁ females will be "super female" but only five out of the 54 females were reported to be "super female" (Tong *et al.* 2010).

Figure 23: Family sex ratio profile of CSD and PSD species.

(A) The Nile tilapia is a CSD species with male heterogametic sex chromosomes (XY). Huge proportion of the 42 families had sex ratio between 41% to 60% males. Data was obtained from (Baroiller *et al.* 2009). (B) The Florida red tilapia is a hybrid species that utilizes PSD system. The sex ratio profile of 80 families showed severely sex-biased broods that were not observed in the CSD species. Data was obtained from (Desprez *et al.* 2006).

In an interesting development with regards to zebrafish SD, Postlethwait's group recently reported data suggesting that zebrafish chromosome 4 might be an "aging" sex chromosome, which had lost its dominant sex determining function (Anderson *et al.* 2012). They found that the long arm of chromosome 4 was late replicating, highly heterochromatic and repeat-rich, which are characteristics of a classical sex chromosome. In addition, out of the two loci they identified with significant linkage to sex by genome-wide linkage study, one of them was a 2 Mb region (60,837,953 to 61,176,889 bp on Zv9) on chromosome 4 long arm (Anderson *et al.* 2012). Hence, these results led them to suggest that chromosome 4 used to be a sex chromosome, however its sex determining function might have been diminished by other stronger sex modifiers located on the autosomes. In view of our aCGH results which did not identify any universal sex-linked difference in chromosome 4 (see section 0), the probes distribution on the higher density HX3 aCGH microarray was checked to attempt to find out the reason. A probe gap of 572 kb was found in the chromosome 4 long arm (61,453,350 to 62,025,960 bp). This is the largest aCGH probe gap on the entire chromosome 4 which also overlapped the 2 Mb sex-linked region reported by Postlethwait's group. The probe gap supported their observation that this region was repeat-rich as the aCGH overall design strategy was to avoid genomic regions with high repeat contents which may cause problems during analysis. Although from our aCGH probes distribution it seems to be in agreement with the "aging" sex chromosome hypothesis, this could not be used to confirm the hypothesis. Nevertheless, if autosomal factors can override the weak sex chromosome function the system cannot be considered as CSD. This is the case in zebrafish as the sex-associated locus on chromosome 4 could not account for all the variance of the sex trait in

their zebrafish population (Anderson *et al.* 2012) and this locus was not found to be associated with sex in two other strains (Bradley *et al.* 2011; Howe *et al.* 2013).

4.3 The zebrafish has a PSD system

Data obtained from different experiments performed for this project and results from other published studies indicated that zebrafish is not a CSD species. The other type of GSD system is PSD in which the cumulative effect of multiple autosomal factors determines the sex (Kosswig 1964; Moore and Roberts 2013). According to Bull (Bull 1983), the characteristics of PSD system are: 1) wide range of progeny sex ratios among different families; 2) strong influence from parental genotypes; and 3) sex ratio shift in response to selection. All these three characteristics of the PSD system were observed in zebrafish.

The first characteristic of wide-ranging progeny sex ratios was observed in the 62 progeny sex ratios analysed for this project (Figure 5). It was also observed in the repeated pairwise mating (see section 3.4) and selective breeding (see section 3.5) experiments.

The second characteristic of strong parental genotype influence was shown by the high similarity of progeny sex ratios between broods from the same breeding pair (see section 3.4). Strong parental genotype influence was also observed by Hörstgen-Schwark's group when they were investigating the effect of heat induced masculinization in zebrafish (Abozaid *et al.* 2011, 2012). They observed that offspring from gynogenetic males showed stronger effect of heat-induced masculinization than progenies sired by wild-type males. In addition, they also

found that progeny sex ratios from the same breeding pairs were highly repeatable (Abozaid *et al.* 2011).

The third characteristic of sex ratio response to selection was observed in the selective breeding experiment carried out to produce sex-biased broods (see section 3.5). The ability to reduce progeny sex ratio variation and an all-male family obtained after four generations of selective breeding fulfilled the third characteristic.

In addition to our results, data reported by other groups also conform to the PSD system. Direct molecular evidences were produced by three genome-wide linkage studies (Bradley *et al.* 2011; Anderson *et al.* 2012; Howe *et al.* 2013) that were published during the course of this project. A total of four loci on chromosome 3, 4, 5 and 16 were found to have significant linkage to sex in five families of zebrafish tested in the three independent studies. At the moment it is unclear what are the genes that contribute to SD in zebrafish and further investigation is required. Therefore, we conclude that the type of GSD mechanism primarily used by zebrafish is PSD.

4.4 A concept for variable sex ratios observed in PSD species

In the PSD system, the sex of an individual is not determined by the presence or absence of a special chromosome. Instead, a decision is made by the allelic combinations of multiple loci, whose products are involved in the SD process. These loci are typically dispersed throughout the genome, but in some species a

few of them might be located on a preferential pair of chromosomes. In other words, the presence of a sex chromosome, per se, should not exclude PSD.

The complexity of a PSD system is the result of different allelic combinations with variable effect on gonad differentiation. The concept is better explained in a simplified PSD model where protein products of four genes located on different chromosomes are involved in SD. Assuming that two of the protein products perform pro-female function that pushes the gonad towards femaleness, while the remaining two are proteins with pro-male function. To further complicate matters, different alleles of the same genes have strong or weak effect. If we assign a binary code to the allele effect (strong = 1 and weak = 0), then the sex of this imaginary PSD species could be predicted by simply comparing the ratio of male to female alleles. Gonad differentiation will be directed to the sex with highest proportion of strong alleles (Figure 24A & B).

Figure 24: A simplified mechanistic model for PSD system based on the involvement of four autosomal genes.

The allelic sets of a (A) male and (B) female brooder. (C&D) Genotypes for two out of the possible 16 allelic combinations in the sperm. Please note that the sperm depicted under (C) carries more strong pro-male alleles while the other sperm (D) has two weak pro-male alleles that can be overridden by a strong pro-female allele. (E&F) Genotypes for two out of the possible 16 allelic combinations in the oocyte. Please note that the oocyte depicted under (E) carries two strong pro-female alleles while the other oocyte (F) has two weak pro-female alleles that can be overridden by a strong pro-male allele. Symbols with different shapes label genes located on four different autosomes. Genes with pro-male function are light blue, whereas those with pro-female function are pink. Strong alleles are represented by larger shape and upper case letters while weak alleles have smaller shape and lower case letters. Ratios of strong male to strong female alleles are indicated in brackets [Figure modified from Liew and Orbán (2014)].

In this system, the gametes of PSD species will have a pre-disposition to maleness or femaleness depending on the combination of alleles they contain, regardless if they are spermatozoa or oocytes. For example in our imaginary PSD species, the male has two strong male alleles and one strong female allele (2:1) which in theory could produce sperm cells of 16 different sexual genotypes (Figure 24A), among them some would have excess of pro-male (Figure 24C) or pro-female (Figure 24D) alleles. Similarly, a female with one strong pro-male and two strong pro-female alleles (1:2; Figure 24B) could also produce oocytes of 16 sexual genotypes and similarly some of them would have more pro-female (Figure 24E) or pro-male (Figure 24F) alleles. Using these highly diverse gametes to produce the next generation will result in similar complex sexual genotypes in the offspring. This phenomenon was observed in a study conducted by Bradley and colleagues, who genotyped two sex-linked loci in a group of full-sib zebrafish and found sex ratios associated with each genotype to be highly variable [see Supplementary Table 4 of Bradley *et al.* (2011)]. Only one genotype produced an all-male phenotype, while the other allelic combinations had different proportions of males and females (Figure 25). This variable sex ratio associated with different genotypes indicates that the PSD system of zebrafish likely involves more than two loci. This is in sharp contrast with the gametes of CSD species, which have only two types of genetic varieties.

Figure 25: Influence of zebrafish parental ‘SD allele sets’ on offspring sex indicates multiple loci are involved.

Genotypes and sex ratios data from Bradley *et al.* (2011) were used to construct this figure [see Supplementary Table 4 in Bradley *et al.* (2011)]. (A) A male and female zebrafish which had heterozygous alleles for the two sex-linked loci on chromosome 5 (blue cube) and chromosome 16 (red hexagon) were crossed. (B) Different proportion of males and females associated with each genotype is a strong indication of PSD. Bar chart: male - light blue bar and female - pink bar.

4.5 Zebrafish gonad differentiation starts before the juvenile ovary phase

The data described so far revealed that the primary SD mechanism of zebrafish is PSD, but the time point (or period) when sex is determined is still not known. There had been two reports, which speculated on the SD timing and when put together cover a wide time period from 12 to 30 dpf (Rodríguez-Marí *et al.* 2010; Tong *et al.* 2010). Using transgenic zebrafish (Wang *et al.* 2007) and gonadal histology study (Maack and Segner 2003) the earliest time point, at which sexually dimorphic phenotypes can be observed could be placed between 16 to 28 dpf (Maack and Segner 2003; Wang *et al.* 2007). To sum it up, these data seem to suggest that zebrafish SD starts around 12 dpf and could stretch over an extensive period which overlaps with gonad differentiation. However, based on data from studying the transgenic *Tg(vasa:vasa-EGFP)* zebrafish line during the early developmental period and results of morpholino-based knockdown experiments, my colleagues postulated that zebrafish SD begins at an earlier period between fertilization and 10 dpf [review: (Orban *et al.* 2009)]. Hence, one of the objectives for this project was to try to confirm - and if possible specify - the SD period using comparative transcriptomic analysis. This transcriptomic approach leverages on the major restructuring of the bipotential gonad into testis or ovary resulting in distinct transcriptome changes during the gonadal differentiation process that follows immediately after SD. Therefore, the earliest time point when gonadal transcriptome starts to bifurcate is an indication for the inception of SD.

The dissection of gonadal tissue from the tiny zebrafish larvae (total length ca. 4 mm) is technically challenging. In order to overcome this technical limit, we used

eviscerated trunks with intact gonads for comparative transcriptomic studies (see section 2.8 for technical details). To further minimize “noise” from other tissues (e.g. muscle) we profiled a selected set of 20 genes from either the microarray experiment (section 3.6) or sex candidate genes from published reports that showed sexually dimorphic expression in the gonad (see Table 7 for complete list of genes). This method was tested on the 28 dpf trunk samples as it was the earliest time point that visual sexing of the zebrafish based on their gonadal vasa-EGFP fluorescence can be performed with high confidence. Groups created by molecular and visual methods were in good agreement with the former having higher sensitivity, as it was possible to sort those zebrafish visually classified as intermediate fluorescence (see section 3.8). As vasa-EGFP intensity can be used to predict the future sex of juvenile zebrafish (Wang *et al.* 2007), this shows that gene expression profiling method could also be used to sex zebrafish larvae and juveniles.

Interestingly, three expression profiles were observed in the 14 to 35 dpf zebrafish instead of the expected one group each for the weak (future male) and strong (future female) fluorescence zebrafish as in 42 dpf. For the 28 (Figure 19C & D) and 35 (Figure 19E & F) dpf juvenile zebrafish, two groups were formed among the weak fluorescence fish in addition to a single group consisting of only the strong fluorescence fish. This variation observed in weak fluorescence zebrafish were most likely due to the different rate of gonadal transformation from juvenile ovary to testis and is in agreement to previous observation of males having different extents of the juvenile ovary phase (Wang *et al.* 2007). Another unexpected result was the sorting of the 7 dpf larvae into two groups (Figure 22 E

& F) instead. The lack of a third group at 7 dpf is most probably due to reduced qPCR array sensitivity caused by dual effects of lesser genes with valid data and diluted expression level because of small sample size. This can be improved by analyzing more genes and more samples in the future.

Despite the shortcomings, combining subtle differences from the seven chosen genes within each 7 dpf individual allowed for clustering of them into different groups. The results clearly show that the gonadal transcriptome had already started to diverge as early as 7 dpf. This is earlier than previous reports by Chung's group who suggested 10 to 12 dpf (Tong *et al.* 2010) and Postlethwait's group who postulated 21 to 30 dpf (Rodríguez-Marí *et al.* 2010). Reason for both groups to find a timing later than our result could be due to the fact that they compared expression level of only one or three sex candidate genes (i.e. *vasa*, *amh* and *cyp19a1a*) by *in situ* hybridization. Although these sex candidate genes are well characterized and known to have sexually dimorphic expression, they are not MSR genes in zebrafish and their sex-specific expression patterns could be a late consequence of SD. Furthermore, estimation of expression levels by *in situ* hybridization will overlook samples with low or near equal expression, unlike quantitative PCR, which is more sensitive.

Besides specifying the SD timing of zebrafish, we also found that the use of trunk samples allowed fast and efficient analysis of gonadal transcriptomes from multiple zebrafish larvae, which would not be easy with other techniques, such as laser microdissection. The use of trunk samples for such study was further supported by a recent published report, which showed that the expression patterns

of sex candidate genes are very similar between Nile tilapia larvae trunk and gonad samples (Poonlaphdecha *et al.* 2013). Careful selection of genes for analysis also increased the sensitivity of molecular sexing by gene expression profiling.

4.6 Summary and future outlook

Before this project, SD was a poorly studied area of zebrafish developmental biology. The type of SD mechanism used by zebrafish was investigated in this project using classical breeding experiments (see sections 3.1, 3.4 and 3.5) and newly developed molecular methods (see sections 3.2 and 0). Results from these experiments showed convincingly that the primary SD mechanism of zebrafish is GSD with characteristics of a PSD species (see section 4.3) and refuted the possibility of a CSD system. This is in agreement with observations by other groups who performed whole genome manipulation studies (Table 1) or analyzed progeny sex ratios produced by breeding of sex-reversed zebrafish (see section 1.4.3). Molecular evidence confirming our result that zebrafish is a PSD species was obtained recently when three genome-wide linkage studies (Bradley *et al.* 2011; Anderson *et al.* 2012; Howe *et al.* 2013) reported a total of four autosomal loci linked to the sex trait in five zebrafish families. Several studies had observed various environmental factors affecting zebrafish sex ratio, albeit the influences were either minor or dependent on the fish genotype (see section 1.5). In fact, interaction between genotype and environment (GxE interaction) is one of the characteristics of a PSD system (Bulmer and Bull 1982; Bull 1983).

After finding out that the primary mode of SD for zebrafish is PSD, we then tried to specify the time period when SD starts. This was done by performing

comparative transcriptomic analysis on the developing gonads using a custom-designed expression microarray (see section 3.6) and a high throughput qPCR array (see section 3.9). The aim of this transcriptomic approach was to identify the earliest time point that the gonadal transcriptomes bifurcate, a sign that is closely related to the SD timing. In this project signs of zebrafish gonadal transcriptome divergence were observed as early as 7 dpf (Figure 22E & F) indicating that SD had already started before that time point. In a similar small-scale qPCR analysis by another group, sexually dimorphic expression of *dmrt1* was observed at around 10 dpf (Jorgensen *et al.* 2008). Both our's and Jorgensen's qPCR results indicated that zebrafish SD occurs earlier than the previously suggested 12 to 30 dpf (Rodríguez-Marí *et al.* 2010; Tong *et al.* 2010) and before the onset of juvenile ovary phase [around 17 to 18 dpf (Takahashi 1977; Maack and Segner 2003)].

During the course of this project, in addition to deciphering the sexual developmental process of zebrafish, two molecular tools each for genomic (i.e. FluoMEP and aCGH microarray) and transcriptomic (i.e. expression microarray and qPCR sexing array) analysis were also developed in collaboration with other lab members. The usefulness of these tools have been shown here and they can also be used for other fields of zebrafish research especially the genomic tools which are currently lacking from the research area and commercial market.

The finding of a PSD system in zebrafish has advanced our knowledge on its sexual development, but at the same time it has also revealed that the gonad developmental process is more complex than it was previously thought. In a PSD system, sex is determined by the interactions of several genetic factors/pathways

that have either mutually supportive or antagonistic relationships with added complexity from minor environmental factors influence. Therefore, we hypothesize that multiple layers of regulations are required to ensure that this intricate network of interactions operate seamlessly to achieve the desirable gonadal phenotypes (i.e. testis or ovary). Knowledge in this area is still lacking due to current technical limits for studying polygenic traits. However, the advancement of high throughput molecular analysis platforms, such as next generation sequencing, can be of great help to decipher the processes. One layer of regulation that has benefited from this advancement is epigenetic control of SD and gonad differentiation [review: (Piferrer 2013)]. Recently it was discovered that zebrafish inherited its DNA methylome from the sperm and not the oocyte (Jiang *et al.* 2013; Potok *et al.* 2013), which might have implication on its sexual development. The microRNA (miRNA)-based regulation is another area that has seen great advancement lately with the discovery of new zebrafish miRNAs using next generation sequencing (Soares *et al.* 2009). Many developmental processes are known to be regulated by miRNA with some linked to sexual development, such as miR-430 mediated migration of zebrafish primordial germ cell (Staton *et al.* 2011). Germ cells are important for zebrafish gonad development, as a bipotential gonad, which lacks germ cells develops an empty testis-like ‘shell’ (Slanchev *et al.* 2005; Siegfried and Nusslein-Volhard 2008) and a mature ovary when depleted of germ cells converts readily to testicular phenotype (Dranow *et al.* 2013).

The complexity of a polygenic trait requires an experimental strategy that is different from those used to analyze the more familiar and easier monogenic traits. These strategies should be designed by drawing from the decades of experience in

studying human polygenic disorders [review: (O'Brien *et al.* 2000; Jones *et al.* 2013)] in combination with the latest advancement in large scale molecular studies. For example, genome wide association studies have benefited from the advancement of high throughput genomic analytical tools (e.g. SNP genotyping and whole genome resequencing) that allow analysis of a large numbers of individual genomic variations (Womack *et al.* 2012). The application of these new research strategies to zebrafish should be easier than to mammalian model organisms due to the higher fecundity, faster generation time and relatively easy maintenance of the former. In the same spirit, any success in the study of PSD system in zebrafish could possibly be extrapolated to the research of mammalian polygenic and sexual disorders in the future. Other interesting questions that arise from the result of this project are: 1) what additional pathways are involved and 2) why does the bipotential gonad of future males still go through the juvenile ovary phase even though differentiation process has started earlier?

5 References

Abozaid H., Wessels S. & Horstgen-Schwark G. (2011) Effect of rearing temperatures during embryonic development on the phenotypic sex in zebrafish (*Danio rerio*). *Sex Dev* **5**: 259-265.

Abozaid H., Wessels S. & Horstgen-Schwark G. (2012) Elevated temperature applied during gonadal transformation leads to male bias in zebrafish (*Danio rerio*). *Sex Dev* **6**: 201-209.

Abucay J.S., Mair G.C., Skibinski D.O.F. & Beardmore J.A. (1999) Environmental sex determination: the effect of temperature and salinity on sex ratio in *Oreochromis niloticus* L. *Aquaculture* **173**: 219-234.

Amores A. & Postlethwait J.H. (1998) Banded chromosomes and the zebrafish karyotype. in *The zebrafish: Genetics and genomics* (eds. Detrich H.W., Westerfield M. & Zon L.I.), pp. 323-338. Academic Press, San Diego, CA.

Anderson J.L., Rodriguez Mari A., Braasch I., Amores A., Hohenlohe P., Batzel P. & Postlethwait J.H. (2012) Multiple sex-associated regions and a putative sex chromosome in zebrafish revealed by RAD mapping and population genomics. *PLoS One* **7**: e40701.

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

Arkhipchuk V.V. (1995) Role of chromosomal and genome mutations in the evolution of bony fishes. *Hydrobiol J* **31**: 55-65.

Ashburner M., Ball C.A., Blake J.A., Botstein D., Butler H., Cherry J.M., Davis A.P., Dolinski K., Dwight S.S., Eppig J.T. *et al.* (2000) Gene Ontology: tool for the unification of biology. *Nat Genet* **25**: 25-29.

Ayers K.L., Smith C.A. & Lambeth L.S. (2013) The molecular genetics of avian sex determination and its manipulation. *Genesis* **51**: 325-336.

Baroiller J.F., D'Cotta H., Bezault E., Wessels S. & Hoerstgen-Schwark G. (2009) Tilapia sex determination: Where temperature and genetics meet. *Comp Biochem Physiol, A: Mol Integr Physiol* **153**: 30-38.

Beer R.L. & Draper B.W. (2013) *nanos3* maintains germline stem cells and expression of the conserved germline stem cell gene *nanos2* in the zebrafish ovary. *Dev Biol* **374**: 308-318.

Benjamini Y. & Hochberg Y. (1995) Controlling the false discovery rate: A practical and powerful approach to multiple testing. *J Roy Stat Soc B Met* **57**: 289-300.

Bernard P. & Harley V.R. (2007) *Wnt4* action in gonadal development and sex determination. *Int J Biochem Cell B* **39**: 31-43.

Bowles J., Feng C.W., Spiller C., Davidson T.L., Jackson A. & Koopman P. (2010) FGF9 suppresses meiosis and promotes male germ cell fate in mice. *Dev Cell* **19**: 440-449.

Bowles J., Knight D., Smith C., Wilhelm D., Richman J., Mamiya S., Yashiro K., Chawengsaksophak K., Wilson M.J., Rossant J. *et al.* (2006) Retinoid signaling determines germ cell fate in mice. *Science* **312**: 596-600.

Bradley K.M., Breyer J.P., Melville D.B., Broman K.W., Knapik E.W. & Smith J.R. (2011) An SNP-based linkage map for zebrafish reveals sex determination loci. *G3* **1**: 3-9.

Brazma A., Hingamp P., Quackenbush J., Sherlock G., Spellman P., Stoeckert C., Aach J., Ansorge W., Ball C.A., Causton H.C. *et al.* (2001) Minimum information about a microarray experiment (MIAME) - toward standards for microarray data. *Nat Genet* **29**: 365-371.

Bridges C.B. (1921) Triploid intersexes in *Drosophila melanogaster*. *Science* **54**: 252-254.

Bull J.J. (1983) *Evolution of sex determining mechanisms*. Benjamin-Cummings Publishing Company, Melon Park, CA.

Bulmer M.G. & Bull J.J. (1982) Models of polygenic sex determination and sex-ratio control. *Evolution* **36**: 13-26.

Chang A., Liew W.C., Chuah A., Lim Z., Lin Q. & Orban L. (2007) FluoMEP: a new genotyping method combining the advantages of randomly amplified polymorphic DNA and amplified fragment length polymorphism. *Electrophoresis* **28**: 525-534.

Charnov E.L. & Bull J.J. (1977) When is sex environmentally determined? *Nature* **266**: 828-830.

Chen M., Zou M., Yang L. & He S. (2012) Basal jawed vertebrate phylogenomics using transcriptomic data from Solexa sequencing. *PLoS One* **7**: e36256.

Chen T.R. & Reisman H.M. (1970) A comparative chromosome study of the North American species of sticklebacks (Teleostei:Gasterosteidae). *Cytogenetics* **9**: 321-332.

Chen W. & Ge W. (2013) Gonad differentiation and puberty onset in the zebrafish: Evidence for the dependence of puberty onset on body growth but not age in females. *Mol Reprod Dev* **80**: 384-392.

Cline T.W. (1993) The *Drosophila* sex determination signal: How do flies count to two? *Trends Genet* **9**: 385-390.

Corley-Smith G.E., Lim C.J. & Brandhorst B.P. (1996) Production of androgenetic zebrafish (*Danio rerio*). *Genetics* **142**: 1265-1276.

Daga R., Thode G. & Amores A. (1996) Chromosome complement, C-banding, Ag-NOR and replication banding in the zebrafish *Danio reio*. *Chromosome Res* **4**: 29-32.

Danzmann R.G., Cairney M., Davidson W.S., Ferguson M.M., Gharbi K., Guyomard R., Holm L.E., Leder E., Okamoto N., Ozaki A. *et al.* (2005) A comparative analysis of the rainbow trout genome with 2 other species of fish (Arctic charr and Atlantic salmon) within the tetraploid derivative Salmonidae family (subfamily: Salmoninae). *Genome* **48**: 1037-1051.

de Almeida Toledo L.F. & Foresti F. (2001) Morphologically differentiated sex chromosomes in neotropical freshwater fish. *Genetica* **111**: 91-100.

Delbridge M. & Graves J. (1999) Mammalian Y chromosome evolution and the male-specific functions of Y chromosome-borne genes. *Rev Reprod* **4**: 101-109.

Desprez D., Briand C., Hoareau M.C., Melard C., Bosc P. & Baroiller J.F. (2006) Study of sex ratio in progeny of a complex *Oreochromis* hybrid, the Florida red tilapia. *Aquaculture* **251**: 231-237.

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

Dranow D.B., Tucker R.P. & Draper B.W. (2013) Germ cells are required to maintain a stable sexual phenotype in adult zebrafish. *Dev Biol* **376**: 43-50.

Draper B.W., McCallum C.M. & Moens C.B. (2007) *nanos1* is required to maintain oocyte production in adult zebrafish. *Dev Biol* **305**: 589-598.

Edgar R.C. (2010) Search and clustering orders of magnitude faster than BLAST. *Bioinformatics* **26**: 2460-2461.

Eimon P. & Ashkenazi A. (2010) The zebrafish as a model organism for the study of apoptosis. *Apoptosis* **15**: 331-349.

Ellegren H. (2011) Sex-chromosome evolution: recent progress and the influence of male and female heterogamety. *Nat Rev Genet* **12**: 157-166.

Endo A. & Ingalls T.H. (1968) Chromosomes of the zebra fish. A model for cytogenetic, embryologic, and ecologic study. *J Hered* **59**: 382-384.

Engeszer R.E., Patterson L.B., Rao A.A. & Parichy D.M. (2007) Zebrafish in the wild: a review of natural history and new notes from the field. *Zebrafish* **4**: 21-40.

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

Fontana F., Chiarelli B. & Rossi A.C. (1970) Il cariotipo di alcune specie di Cyprinidae, Centrarchidae, Characidae studiate mediante culture *in vivo*. *Caryologia* **23**: 549-564.

Froese R. & Pauly D. (2013) *FishBase*. Accessed on 20 Jun 2013, from www.fishbase.org

Gornung E., Gabrielli I., Cataudella S. & Sola L. (1997) CMA3-banding pattern and fluorescence *in situ* hybridization with 18S rRNA genes in zebrafish chromosomes. *Chromosome Res* **5**: 40-46.

Graves J.A.M. (1995) The origin and function of the mammalian Y chromosome and Y-borne genes--an evolving understanding. *Bioessays* **17**: 311-320.

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

Hayden M.S. & Ghosh S. (2008) Shared principles in NF- κ B signaling. *Cell* **132**: 344-362.

Hazlerigg C.R., Lorenzen K., Thorbek P., Wheeler J.R. & Tyler C.R. (2012) Density-dependent processes in the life history of fishes: evidence from laboratory populations of zebrafish *Danio rerio*. *PLoS One* **7**: e37550.

Hoebeeck J., Speleman F. & Vandesompele J. (2007) Real-time quantitative PCR as an alternative to southern blot or fluorescence *in situ* hybridization for detection of gene copy number changes. in *Protocols for nucleic acid analysis by nonradioactive probes* (eds. Hilario E. & Mackay J.), pp. 205-226. Humana Press, New York, NY.

Hörstgen-Schwark G. (1993) Production of homozygous diploid zebra fish (*Brachydanio rerio*). *Aquaculture* **112**: 25-37.

Horváth L. & Orbán L. (1995) Genome and gene manipulation in the common carp. *Aquaculture* **129**: 157-181.

Howe K., Clark M.D., Torroja C.F., Torrance J., Berthelot C., Muffato M., Collins J.E., Humphray S., McLaren K., Matthews L. *et al.* (2013) The zebrafish reference genome sequence and its relationship to the human genome. *Nature* **496**: 498-503.

Ihssen P.E., McKay L.R., Mcmillan I. & Phillips R.B. (1990) Ploidy manipulation and gynogenesis in fishes - cytogenetic and fisheries applications. *Trans Am Fish Soc* **119**: 698-717.

Irizarry R.A., Hobbs B., Collin F., Beazer-Barclay Y.D., Antonellis K.J., Scherf U. & Speed T.P. (2003) Exploration, normalization, and summaries of high density oligonucleotide array probe level data. *Biostatistics* **4**: 249-264.

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

Jones M., Fejes A. & Kobor M. (2013) DNA methylation, genotype and gene expression: Who is driving and who is along for the ride? *Genome Biol* **14**: 126.

Jorgensen A., Morthorst J.E., Andersen O., Rasmussen L.J. & Bjerregaard P. (2008) Expression profiles for six zebrafish genes during gonadal sex differentiation. *Reprod Biol Endocrinol* **6**: 25.

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

Kanehisa M. & Goto S. (2000) KEGG: Kyoto encyclopedia of genes and genomes. *Nucleic Acids Res* **28**: 27-30.

Kirankumar S., Anathy V. & Pandian T.J. (2003) Hormonal induction of supermale golden rosy barb and isolation of Y-chromosome specific markers. *Gen Comp Endocrinol* **134**: 62-71.

Kirankumar S. & Pandian T.J. (2004) Production and progeny testing of androgenetic rosy barb *Puntius conchoni*. *J Exp Zool Part A* **301A**: 938-951.

Kochakpour N. & Moens P.B. (2008) Sex-specific crossover patterns in zebrafish (*Danio rerio*). *Heredity* **100**: 489-495.

Komen H. & Thorgaard G.H. (2007) Androgenesis, gynogenesis and the production of clones in fishes: A review. *Aquaculture* **269**: 150-173.

Kosswig C. (1964) Polygenic sex determination. *Experientia* **20**: 190-199.

Koubova J., Menke D.B., Zhou Q., Capel B., Griswold M.D. & Page D.C. (2006) Retinoic acid regulates sex-specific timing of meiotic initiation in mice. *Proc Natl Acad Sci USA* **103**: 2474-2479.

Krøvel A.V. & Olsen L.C. (2002) Expression of a *vas::EGFP* transgene in primordial germ cells of the zebrafish. *Mech Develop* **116**: 141-150.

Lane D.P. & Crawford L.V. (1979) T antigen is bound to a host protein in SY40-transformed cells. *Nature* **278**: 261-263.

Lawrence C., Ebersole J.P. & Kesseli R.V. (2008) Rapid growth and out-crossing promote female development in zebrafish (*Danio rerio*). *Environ Biol Fishes* **81**: 239-246.

Li C., Orti G., Zhang G. & Lu G. (2007) A practical approach to phylogenomics: the phylogeny of ray-finned fish (Actinopterygii) as a case study. *BMC Evol Biol* **7**: 44.

Lieschke G.J. & Currie P.D. (2007) Animal models of human disease: zebrafish swim into view. *Nat Rev Genet* **8**: 353-367.

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

Liew W.C. & Orbán L. (2014) Zebrafish sex: a complicated affair. *Brief Funct Genomics* **13**: 172-187.

Maack G. & Segner H. (2003) Morphological development of the gonads in zebrafish. *J Fish Biol* **62**: 895-906.

Mank J.E. & Avise J.C. (2009) Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sex Dev* **3**: 60-67.

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

Martinovic D., Villeneuve D.L., Kahl M.D., Blake L.S., Brodin J.D. & Ankley G.T. (2009) Hypoxia alters gene expression in the gonads of zebrafish (*Danio rerio*). *Aquat Toxicol* **95**: 258-272.

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

Matsuda M., Sotoyama S., Hamaguchi S. & Sakaizumi M. (1999) Male-specific restriction of recombination frequency in the sex chromosomes of the medaka, *Oryzias latipes*. *Genet Res* **73**: 225-231.

McLaren A. (1988) Sex determination in mammals. *Trends Genet* **4**: 153-157.

Meyer A. & Van de Peer Y. (2005) From 2R to 3R: evidence for a fish-specific genome duplication (FSGD). *Bioessays* **27**: 937-945.

Michelmore R.W., Paran I. & Kesseli R.V. (1991) Identification of markers linked to disease-resistance genes by bulked segregant analysis: A rapid method to detect markers in specific genomic regions by using segregating populations. *Proc Natl Acad Sci USA* **88**: 9828-9832.

Modi W.S. & Crews D. (2005) Sex chromosomes and sex determination in reptiles. *Curr Opin Genet Dev* **15**: 660-665.

Moore E.C. & Roberts R.B. (2013) Polygenic sex determination. *Curr Biol* **23**: R510-R512.

Myosho T., Otake H., Masuyama H., Matsuda M., Kuroki Y., Fujiyama A., Naruse K., Hamaguchi S. & Sakaizumi M. (2012) Tracing the emergence of a novel sex-determining gene in Medaka, *Oryzias luzonensis*. *Genetics* **191**: 163-170.

Nanda I., Feichtinger W., Schmid M., Schroder J.H., Zischler H. & Epplen J.T. (1990) Simple repetitive sequences are associated with differentiation of the sex-chromosomes in the guppy fish. *J Mol Evol* **30**: 456-462.

Navarro-Martin L., Vinas J., Ribas L., Diaz N., Gutierrez A., Di Croce L. & Piferrer F. (2011) DNA methylation of the gonadal aromatase (*cyp19a*) promoter is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLoS Genet* **7**: e1002447.

Nef S. & Vassalli J.D. (2009) Complementary pathways in mammalian female sex determination. *J Biol* **8**: 74.

Nelson J.S. (2007) *Fishes of the world, 4th Edition*. John Wiley & Sons, Inc., Hoboken, NJ.

O'Brien S.J., Nelson G.W., Winkler C.A. & Smith M.W. (2000) Polygenic and multifactorial disease gene association in man: Lessons from AIDS1. *Annu Rev Genet* **34**: 563-591.

Ohno S. (1999) The one-to-four rule and paralogues of sex-determining genes. *Cell Mol Life Sci* **55**: 824-830.

Orban L., Sreenivasan R. & Olsson P.E. (2009) Long and winding roads: testis differentiation in zebrafish. *Mol Cell Endocrinol* **312**: 35-41.

Ospina-Alvarez N. & Piferrer F. (2008) Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. *PLoS One* **3**: e2837.

Pandian T.J. & Sheela S.G. (1995) Hormonal induction of sex reversal in fish. *Aquaculture* **138**: 1-22.

Parsons J.E. & Thorgaard G.H. (1985) Production of androgenetic diploid rainbow trout. *J Hered* **76**: 177-181.

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

Pelegri F. & Schulte-Merker S. (1999) A gynogenesis-based screen for maternal-effect genes in the zebrafish *Danio rerio*. in *Method Cell Biol* (eds. Detrich H.W., Westerfield M. & Zon L.I.), pp. 1-20. Academic Press, San Diego, CA.

Phillips R.B. & Reed K.M. (2000) Localization of repetitive DNAs to zebrafish (*Danio rerio*) chromosomes by fluorescence *in situ* hybridization (FISH). *Chromosome Res* **8**: 27-35.

Pieau C., Dorizzi M. & Richard-Mercier N. (1999) Temperature-dependent sex determination and gonadal differentiation in reptiles. *Cell Mol Life Sci* **55**: 887-900.

Piferrer F. (2013) Epigenetics of sex determination and gonadogenesis. *Dev Dyn* **242**: 360-370.

Piferrer F. & Guiguen Y. (2008) Fish gonadogenesis. Part II: Molecular biology and genomics of sex differentiation. *Rev Fish Sci* **16**: 35 - 55.

Piferrer F., Ribas L. & Diaz N. (2012) Genomic approaches to study genetic and environmental influences on fish sex determination and differentiation. *Mar Biotechnol* **14**: 591-604.

Pijnacker L.P. & Ferwerda M.A. (1995) Zebrafish chromosome banding. *Genome* **38**: 1052-1055.

Poonlaphdecha S., Pepey E., Canonne M., de Verdal H., Baroiller J.F. & D'Cotta H. (2013) Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *dmrt1* and *amh* expressions. *Gen Comp Endocrinol* **193**: 234-242.

Post A. (1964) Vergleichende untersuchungen der chromosomenzahlen bei süsswassernteleostern. *Z Zool Syst Evol* **3**: 47-93.

Postlethwait J.H., Johnson S.L., Midson C.N., Talbot W.S., Gates M., Ballinger E.W., Africa D., Andrews R., Carl T., Eisen J.S. *et al.* (1994) A genetic linkage map for the zebrafish. *Science* **264**: 699-703.

Postlethwait J.H., Yan Y.L., Gates M.A., Horne S., Amores A., Brownlie A., Donovan A., Egan E.S., Force A., Gong Z. *et al.* (1998) Vertebrate genome evolution and the zebrafish gene map. *Nat Genet* **18**: 345-349.

Potok M.E., Nix D.A., Parnell T.J. & Cairns B.R. (2013) Reprogramming the maternal zebrafish genome after fertilization to match the paternal methylation pattern. *Cell* **153**: 759-772.

Pradhan A., Khalaf H., Ochsner S.A., Sreenivasan R., Koskinen J., Karlsson M., Karlsson J., McKenna N.J., Orban L. & Olsson P.E. (2012) Activation of NF-kappaB protein prevents the transition from juvenile ovary to testis and promotes ovarian development in zebrafish. *J Biol Chem* **287**: 37926-37938.

Purdom C.E. (1969) Radiation-induced gynogenesis and androgenesis in fish. *Heredity* **24**: 431-444.

Ravi V. & Venkatesh B. (2008) Rapidly evolving fish genomes and teleost diversity. *Curr Opin Genet Dev* **18**: 544-550.

Rishi K.K. (1976) Karyotypic studies on four species of fishes. *Nucleus* **19**: 95-98.

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

Rodríguez-Marí A. & Postlethwait J.H. (2011) The role of fanconi anemia/BRCA genes in zebrafish sex determination. in *The zebrafish: Disease models and*

chemical screens (eds. Detrich H.W., Westerfield M. & Zon L.I.), pp. 461-490. Academic Press, San Diego, CA.

Rozen S. & Skaletsky H. (1999) Primer3 on the WWW for general users and for biologist programmers. in *Bioinformatics Methods and Protocols* (eds. Misener S. & Krawetz S.), pp. 365-386. Humana Press, New York, NY.

Sambrook J. & Russell D.W. (2006) Purification of nucleic acids by extraction with phenol:chloroform. *CSH Protoc*: pdb.prot4455.

Schreeb K., Groth G., Sachsse W. & Freundt K. (1993) The karyotype of the zebrafish (*Brachydanio rerio*). *J Exp Anim Sci* **36**: 27-31.

Segner H. (2009) Zebrafish (*Danio rerio*) as a model organism for investigating endocrine disruption. *Comp Biochem Physiol, C: Toxicol Pharmacol* **149**: 187-195.

Ser J.R., Roberts R.B. & Kocher T.D. (2010) Multiple interacting loci control sex determination in Lake Malawi cichlid fish. *Evolution* **64**: 486-501.

Shang E.H., Yu R.M. & Wu R.S. (2006) Hypoxia affects sex differentiation and development, leading to a male-dominated population in zebrafish (*Danio rerio*). *Environ Sci Technol* **40**: 3118-3122.

Sharma K., Sharma O. & Tripathi M. (1998) Female heterogamety in *Danio rerio* (Cypriniformes: Cyprinidae). *Proc Indian Natl Sci Acad (B Biol Sci)* **68(B)**: 123-126.

Siegfried K.R. & Nusslein-Volhard C. (2008) Germ line control of female sex determination in zebrafish. *Dev Biol* **324**: 277-287.

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

Singer A., Perlman H., Yan Y., Walker C., Corley-Smith G., Brandhorst B. & Postlethwait J. (2002) Sex-specific recombination rates in zebrafish (*Danio rerio*). *Genetics* **160**: 649-657.

Skalitzky C., Heckel T., Wessner S., Wijdicks M. & Watt M.A. (2011) Roche cDNA synthesis system for use with NimbleGen gene expression microarrays. in *Gene Expression: Technical Note*.

Slanchev K., Stebler J., de la Cueva-Méndez G. & Raz E. (2005) Development without germ cells: The role of the germ line in zebrafish sex differentiation. *Proc Natl Acad Sci USA* **102**: 4074-4079.

Soares A., Pereira P., Santos B., Egas C., Gomes A., Arrais J., Oliveira J., Moura G. & Santos M. (2009) Parallel DNA pyrosequencing unveils new zebrafish microRNAs. *BMC Genomics* **10**: 195.

Spence R., Gerlach G., Lawrence C. & Smith C. (2008) The behaviour and ecology of the zebrafish, *Danio rerio*. *Biol Rev Camb Philos Soc* **83**: 13-34.

Sreenivasan R., Cai M., Bartfai R., Wang X., Christoffels A. & Orban L. (2008) Transcriptomic analyses reveal novel genes with sexually dimorphic expression in the zebrafish gonad and brain. *PLoS One* **3**: e1791.

Sreenivasan R., Jiang J., Wang X., Bartfai R., Kwan H.Y., Christoffels A. & Orban L. (2014) Gonad differentiation in zebrafish is regulated by the canonical Wnt signaling pathway. *Biol Reprod* **90**: 45.

Sreenivasan R., Wang X., Bartfai R., Kwan H.Y., Christoffels A. & Orban L. (2009) Global expression profiling in zebrafish reveals genes with potential roles in sexual differentiation. *Biol Reprod* **78**: 116.

Staton A.A., Knaut H. & Giraldez A.J. (2011) miRNA regulation of Sdf1 chemokine signaling provides genetic robustness to germ cell migration. *Nat Genet* **43**: 204-211.

Streisinger G., Walker C., Dower N., Knauber D. & Singer F. (1981) Production of clones of homozygous diploid zebra fish (*Brachydanio rerio*). *Nature* **291**: 293-296.

Subramanian A., Tamayo P., Mootha V.K., Mukherjee S., Ebert B.L., Gillette M.A., Paulovich A., Pomeroy S.L., Golub T.R., Lander E.S. *et al.* (2005) Gene set enrichment analysis: A knowledge-based approach for interpreting genome-wide expression profiles. *Proc Natl Acad Sci USA* **102**: 15545-15550.

Sukumaran R. (2008) *Computational analysis of genotypes produced by FluoMEP*. Honours thesis, National University of Singapore

Takahashi H. (1977) Juvenile hermaphroditism in the zebrafish, *Brachydanio rerio*. *Bull Fac Fish Hokkaido Univ* **28**: 57-65.

Tease C. & Hulten M.A. (2004) Inter-sex variation in synaptonemal complex lengths largely determine the different recombination rates in male and female germ cells. *Cytogenet Genome Res* **107**: 208-215.

Tessema M., Müller-Belecke A. & Hörstgen-Schwark G. (2006) Effect of rearing temperatures on the sex ratios of *Oreochromis niloticus* populations. *Aquaculture* **258**: 270-277.

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. *Dev Biol* **344**: 849-856.

Traut W. & Winking H. (2001) Meiotic chromosomes and stages of sex chromosome evolution in fish: zebrafish, platyfish and guppy. *Chromosome Res* **9**: 659-672.

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. *J Exp Biol* **205**: 711-718.

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

Ueda T. & Naoi H. (1999) BrdU-4Na-EDTA-Giemsa band karyotypes of 3 small freshwater fish, *Danio rerio*, *Oryzias latipes*, and *Rhodeus ocellatus*. *Genome* **42**: 531.

Uwa H. & Ojima Y. (1981) Detailed and banding karyotype analyses of the medaka, *Oryzias latipes* in cultured cells. *P Jpn Acad B-Phys* **57**: 39-43.

Vandeputte M., Dupont-Nivet M., Chavanne H. & Chatain B. (2007) A polygenic hypothesis for sex determination in the European sea bass *Dicentrarchus labrax*. *Genetics* **176**: 1049-1057.

Wallace B.M.N. & Wallace H. (2003) Synaptonemal complex karyotype of zebrafish. *Heredity* **90**: 136-140.

Wang X.G., Bartfai R., Sleptsova-Freidrich I. & Orban L. (2007) The timing and extent of 'juvenile ovary' phase are highly variable during zebrafish testis differentiation. *J Fish Biol* **70**: 33-44.

Westerfield M. (2000) *The zebrafish book. A guide for the laboratory use of zebrafish (Danio rerio)*. Univ. of Oregon Press, Eugene, OR.

Womack J.E., Jang H.J. & Lee M.O. (2012) Genomics of complex traits. *Ann N Y Acad Sci* **1271**: 33-36.

Yamaguchi T. & Kitano T. (2012) High temperature induces *cyp26b1* mRNA expression and delays meiotic initiation of germ cells by increasing cortisol levels during gonadal sex differentiation in Japanese flounder. *Biochem Biophys Res Commun* **419**: 287-292.

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

Ye J., Coulouris G., Zaretskaya I., Cutcutache I., Rozen S. & Madden T. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**: 134.

Yoshimoto S., Okada E., Umemoto H., Tamura K., Uno Y., Nishida-Umehara C., Matsuda Y., Takamatsu N., Shiba T. & Ito M. (2008) A W-linked DM-domain gene, DM-W, participates in primary ovary development in *Xenopus laevis*. *Proc Natl Acad Sci USA* **105**: 2469-2474.

Zhao S. & Fernald R.D. (2005) Comprehensive algorithm for quantitative real-time polymerase chain reaction. *J Comput Biol* **12**: 1047-1064.

Appendix 1: Primer sequences of sex-linked FluoMEP markers for guppy and rosy barb.

Primer sequence	
<u>Common primer</u>	
C117	NNGGAGTTTGCA
<u>RAPD primer</u>	
OPAL-16	CTTTCGAGGG
OPAO-08	ACTGGCTCTC
OPAR-09	CCAGATGGGG
OPL-20	TGGTGGACCA
OPN-20	GGTGCTCCGT

Appendix 2: The distribution of aCGH probes on each chromosome

Probe distribution on HX12 aCGH microarray

Chromosome	Number of probes	Probe distance	
		Median	Mean
1	4903	10070	12249
2	4928	10068	12228
3	5207	10067	12133
4	4078	10086	15221
5	5990	10069	12631
6	4875	10070	12245
7	6366	10067	12083
8	4770	10069	11771
9	4682	10067	12427
10	3836	10068	12146
11	3799	10070	12278
12	3900	10070	12988
13	4418	10069	12239
14	4157	10071	12921
15	3836	10071	12364
16	4528	10075	12969
17	4323	10067	12490
18	4065	10067	12248
19	4210	10066	11905
20	4670	10067	11976
21	3638	10071	12230
22	3151	10076	13372
23	3807	10066	12185
24	3505	10066	12535
25	2773	10072	13864
Mitochondria	2	10046	10046
Total	108,417	10,069	12,500

Probe distribution on HX3 aCGH microarray

Chromosome	Number of probes	Probe distance	
		Median	Mean
1	30726	1684	1955
2	29413	1689	2049
3	31310	1687	2018
4	25347	1692	2451
5	36161	1688	2093
6	30031	1686	1990
7	38705	1687	1995
8	28620	1689	1963
9	28300	1687	2058
10	23253	1688	2004
11	23943	1685	1949
12	23420	1688	2165
13	27581	1687	1962
14	25851	1685	2079
15	23378	1689	2029
16	27430	1688	2142
17	26266	1687	2056
18	24694	1688	2017
19	24886	1690	2015
20	27964	1689	2002
21	21903	1689	2032
22	19594	1688	2156
23	23491	1686	1975
24	21351	1687	2058
25	17391	1685	2213
Mitochondria	9	1687	1671
Total	661,018	1,688	2,052

Appendix 3: Primer sequences used for CNVR gDNA-qPCR validation and their genomic location.

Primer	For primer	Rev primer	Product length (bp)	Chr	Start	End
DrSC19	TGGTTTAATCCACCTGACCA	CAATAGCTCACCCAGCAAAA	127	8	26,618,865	26,618,991
DrSC23	TGTTCCCTGCCAACAAAATA	CAAACCTTTGGCAAGCAGAA	105	7	41,134,519	41,134,623
CNVR1	TGCAAACTGAGCAATGTTGA	TTGGAGGCTAATGTGCAGAG	107	8	30,299,989	30,300,095
CNVR2	AAAGCGTGGGGAAAACTCAATCGG	CCGGTTGGTGTGGCTGTCTCG	157	7	2,938,738	2,938,894
CNVR3	TGACTACCCGAGCTTTGGTGAGG	GCGGATTACTGGTCCGTTAATGTGG	183	8	47,652,991	47,653,173

Appendix 4: Gene details and primer sequences used for expression microarray validation and qPCR array

Gene Symbol	Accession Number	Source	Forward primer	Reverse primer	Amplicon length (bp)	Efficiency
Primers for microarray result validation						
<i>lin28a</i>	NM_201091	Microarray	TGGATGTCTTCGTCCATCAG	CTTTTAAAGGTGAATTCCACTGC	88	0.876
<i>qki2</i>	NM_200842	Microarray	CCAGTGGTGTCTTAGGTATGGC	GCTGAAGACTGTGTTAAACCCC	72	0.931
<i>wu:fd16d01</i>	XM_681781	Microarray	CGTCAGTGTACAACAAGAGAAAACA	GCGCATCCAATACTCCTGAT	94	0.958
<i>zgc:162969</i>	NM_001089408	Microarray	GAAACAAGATGGGTTACGTTAGTTT	CGCCAGTCACGTGTGAGAT	76	0.936
Primers for expression profiling using qPCR array						
<i>amh</i>	NM_001007779	Sex candidate gene	ACAACCCGAAGGTCAACCCGC	GTGGCATGTTGGTCAGTTGGCTG	103	0.846
<i>cyp19a1a</i>	NM_131154	Sex candidate gene	GATATTTGCTCAGAGCCATGGA	GCTCTGCCAGCTAAACACT	187	0.919
<i>dmrt1</i>	NM_205628	Sex candidate gene	TGCCCAGGTGGCGTTACGG	CGGGTGATGGCGGTCTGAG	149	1.024
<i>dnd</i>	NM_212795	Sex candidate gene	TCGTGGAAGCTTTTCGGAACCGG	TGTCCTCGACGCGCTTGGAC	84	1.001
<i>foxl2</i>	NM_001045252	Sex candidate gene	AAACACTGGGAAGGTTGCGTGTC	TTGTCCGGCCCCCTTCTCTGG	146	0.906
<i>ift57</i>	NM_001001832	Sex candidate gene	GAGGTGGAGCGAGTTCTTCC	GACAACCCCTTGGCATCTTGC	133	0.914
<i>pou5f1</i>	NM_131112	Sex candidate gene	GGAAGGCACAGTCCGTTCTGC	CGCACTACATCTCTCTCCAGGCC	114	0.952
<i>vtg5</i>	NM_001025189	Sex candidate gene	AAGGCCCTGCCAACTGACCG	TGGGTCCAGTAACCATCGGTATTGC	124	1.044
<i>zp2</i>	NM_131330	Sex candidate gene	CCCACAGTCTCTGTCTCCTGTGC	TTCTTGCGCAACTCTGCTCAC	105	0.993
CO351530	CO351530	Microarray	CCTCTGGGAAGTTGAATGAATC	AAGGCAAACCACTCCTCA	75	0.874
<i>csad</i>	NM_001007348	Microarray	TGCATGTTGACGCCGCATGG	AGAGCACTGCAGTCCCGTCAG	134	0.874
EV559355	EV559355	Microarray	TGCAGACGTGCACATTTTT	GATCAGACAGACCTGACCAAGA	114	0.972
<i>gtf3ab</i>	NM_001089544	Microarray	AAAGCCTTTGCTATGGAGGAA	TTGGAGGTCTTCTTGCCTTT	110	1.068
<i>loc100334979</i>	XM_002666960	Microarray	GAGAACGGCTTCAACCAAGAG	TCCGCCACTAAATACCTGCT	60	0.923
<i>mcm3l</i>	NM_001171068	Microarray	GGAACAGCAATGCCGTGGG	AGATCTGCAACTCTGCTCAGATGC	91	1.043
<i>retsatl</i>	NM_001040004	Microarray	CCACCATCAGGTGCAATACA	GCAGACAGCGATCCATAACA	129	0.918
<i>si:ch211-119o8.6</i>	NM_001114740	Microarray	TGGGGAACCTGGAAAGTGAA	TGGTCTCCATCAACATCTGG	123	0.894
<i>snpc1a</i>	NM_200263	Microarray	CCAGGACCTATAATCATTTTGA	CTCAAAAGCCAGTCTTTCGTC	64	0.989
<i>wu:fi34b01</i>	XM_001338062	Microarray	CTGGGAAGACACTGCCAAAG	GCCAACGCATCTGAGAGATT	69	0.916
<i>zgc:66449</i>	NM_199801	Microarray	CGCGGTCTGGAGTATTTCTG	CGTCCACTAGGATGGAAATGA	74	1.091
<i>eef1a1l1</i>	NM_131263	Reference gene	GGAGGCTGCCAACTTCAACGC	GGCGATGTGAGCAGTGTGGC	103	1.014
<i>rpl13a</i>	NM_212784	Reference gene	TCTGGAGGACTGTAAGAGGTATGC	AGACGCACAATCTTGAGAGCAG	148	1.001

Appendix 5: Family trees and progeny sex ratios of the two zebrafish lines used for selective breeding

Female-biased H3 line

F_1	H3
	4.8% (n=62)

F_2	Fe1	Fe2	Fe3	Fe4
M1	56.9% (n=62)	4.8% (n=361)	28.6% (n=176)	30.9% (n=110)
M2		17.4% (n=23)	4.5% (n=44)	

H3-21				
F_3	Fe1	Fe2	Fe3	Fe4
M1		48.6% (n=140)		46.3% (n=67)
M2	38.5% (n=52)	40.1% (n=22)	28.9% (n=76)	

H3-22			
F_3	Fe1	Fe2	Fe3
M1	22.3% (n=112)	33.4% (n=154)	27.2% (n=151)
M2	55.0% (n=107)	26.9% (n=52)	16.0% (n=119)
M3		45.6% (n=87)	34.1% (n=176)

Male-biased D line

F_1	Fe1	Fe2	Fe3	Fe4	Fe5
M1	88.9% (n=18)				
M2		87.1% (n=62)			
M3			97.3% (n=75)		
M4				51.8% (n=83)	
M5					94.6% (n=74)

D8 sub-line			
F_2	Fe1	Fe2	Fe3
M1	94.0% (n=217)		
M2	88.3% (n=94)	96.5% (n=173)	
M3			91.1% (n=45)

D4 sub-line		
F_2	Fe1	Fe2
M1	87.2% (n=199)	93.5% (n=62)
M2	80.5% (n=128)	91.5% (n=238)

F_3	Fe1	Fe2	Fe3
M1	88.7% (n=53)	78.9% (n=76)	
M2		77.8% (n=27)	70.1% (n=24)
M3	81.8% (n=33)		82.4% (n=34)

F_3	Mass cross
	84.8% (n=112)

F_4	Fe1	Fe2
M1	98.6% (n=68)	91.4% (n=58)
M2	100% (n=55)	97.0% (n=66)

F_4	Fe1	Fe2	Fe3
M1	61.6% (n=86)	83.5% (n=106)	
M2		94.2% (n=55)	
M3	48.9% (n=89)		87.2% (n=47)