Analysing sex determination in farmed fish using Next Generation DNA Sequencing

A thesis presented for the degree of
Doctor of Philosophy

By

Christos Palaiokostas
MSc

INSTITUTE OF AQUACULTURE

UNIVERSITY OF
STIRLING

2013

Declaration

I hereby declare that this thesis has been composed by myself and is the result of my own investigations. It has neither been accepted, nor submitted for any other degree. All sources of information have been duly acknowledged.

Signed:

Date:

List of published Papers in Peer Reviewed Journals

C. Palaiokostas, M. Bekaert, A. Davie, M. E. Cowan, M. Oral, J.B. Taggart, K. Gharbi, B.J. McAndrew, D.J. Penman, H. Migaud. Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics 14:556*.

C. Palaiokostas, M. Bekaert, M.G.Q. Khan, J.B. Taggart, K. Gharbi, B.J. McAndrew, D.J. Penman. Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (*Oreochromis niloticus* L.) using RAD Sequencing. *PLoS ONE 8(7): e68389. doi:10.1371/journal.pone.0068389*.

Presentations in Conferences

C. Palaiokostas, M. Bekaert, M. Cowan, M. Oral, A. Davie, J.B. Taggart, K. Gharbi, B. J. McAndrew, D. J. Penman, H. Migaud. QTL Mapping and Genomic Selection in Aquaculture species using RAD-seq. 3rd PhD Research Conference, 2012, Stirling, UK.

C. Palaiokostas, M.G.Q. Khan, M. Bekaert, J.B. Taggart, K. Gharbi, B.J. McAndrew, D.J. Penman. Identification of sex-linked genetic markers in Nile tilapia and their usage for all-male production through Marker Assisted Selection. AQUA 2012, Prague, Czech Republic.

H. Migaud, M. Bekaert, **C. Palaiokostas**, M. Cowan, A. Davie, J.B. Taggart, K. Gharbi, B.J. McAndrew, D.J. Penman. Analysis of sex determination in the Atlantic halibut (*Hippoglossus hippoglossus*), using sex reversal progeny testing and RAD sequencing. AQUA 2012, Prague, Czech Republic.

C. Palaiokostas, M.G.Q. Khan, M. Bekaert, J.B. Taggart, K. Gharbi, B.J. McAndrew, D. Penman. Linkage mapping of the LG1 sex determination locus in the Nile tilapia: functional polymorphism and a dense SNP map from RAD sequencing. Genetics in Aquaculture XI, 2012, Auburn, USA.

M. Bekaert, **C. Palaiokostas**, M. Cowan, A. Davie, J.B. Taggart, K. Gharbi, B. McAndrew, D. Penman, H. Migaud. Mapping the sex-determining locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. Genetics in Aquaculture XI, 2012, Auburn, USA.

C Palaiokostas, ME Cowan, MGQ Khan, A Pino-Querido, A Davie, M Bekaert, BJ McAndrew, H Migaud, DJ Penman, JB Taggart. Identifying sex-associated markers in farmed fish using RAD sequencing. 3rd UK RAD sequencing meeting, 6-7 September 2011, Edinburgh, UK.

ACKNOWLEDGEMENTS

I would like to take this opportunity to express my sincere gratitude to my supervisors Dr. David Penman and Professor Brendan McAndrew for encouragement and guidance throughout my PhD study. I also thank them for always being willing to address my numerous questions.

During my PhD studies I had the privilege of working with Dr. John Taggart and Dr. Michael Bekaert. I would like to thank both of them for sharing their expertise with me. John's guidance during my lab work and Michaels' expertise in Bioinformatics were most valuable. My thanks to Professor Herve Migaud, Dr. Andrew Davie, Dr. Marc Vandeputte, Dr. Mairi Cowan and Dr. Karim Gharbi for the excellent collaboration during my PhD.

I would like to give my sincere thanks to Stirling University for funding my PhD research.

My sincere thanks also to Mr. Keith Ranson and Mr. Brian Craig for providing all facilities to breed, rear and maintain experimental animals in the Tropical Aquarium System for almost three years. I would also like to thank the previous members of the Genetics research group Dr. Amy Rajaee and Dr. Milton Khan for their kind help. My thanks also to the current members of the Genetics group Munevver, Khalfan and Udin for their friendship.

I would like also to thank my friends Xristoforos, Greta, Carina, Beatrix, Antonis, Joanna, Sean and Steven for their support and friendship during my PhD study.

Finally I would like to thank Mayra, my brother and my parents for their continual moral support during my PhD journey.

Γηράσκω αεί διδασκόμενος- Σόλων

I always learn as I age- Solon

στην οικογένειά μου

to my family

Abstract

The aim of the current thesis was the analysis of the genetics of sex determination of farmed fish with sexual dimorphism, using Next Generation Sequencing. Three different species of farmed fish with sex-determining systems of varying complexity were studied. Both full-sibs and more distantly related specimens of Atlantic halibut (*Hippoglossus hippoglossus*), Nile tilapia (*Oreochromis niloticus*) and European sea bass (*Dicentrarchus labrax*) were used for this study. Application of Restriction-site Associated DNA sequencing (RAD-seq) and double digest Restriction-site Associated DNA sequencing (ddRAD-seq), two related techniques based on next generation sequencing, allowed the identification of thousands of Single Nucleotide Polymorphisms (SNPs; > 3,000) for each of the above species. The first SNP-based genetic maps for the above species were constructed during the current study.

The first evidence concerning the location of the sex-determining region of Atlantic halibut is provided in this study. In the case of Nile tilapia both novel sex-determining regions and fine mapping of the major sex-determining region are presented. In the study of European sea bass evidence concerning the absence of a major sex-determining gene was provided. Indications of putative sex-determining regions in this species are also provided.

The results of the current thesis help to broaden current knowledge concerning sex determination in three important farmed fish. In addition the results of the current thesis have practical applications as well, towards the production of mono-sex stocks of those species for the aquaculture industry.

Table of Contents

List of Tables

List of Figures

Chapter 1
General Introduction

1 General Introduction

1.1 Sex determination in fish

1.1.1 General Review

Sex determination and differentiation in fish varies, with a wide range of gonochoristic and hermaphroditic species (Devlin & Nagahama, 2002). In addition a wide variety of fish species produced in aquaculture exhibit sexual dimorphism in a range of traits of interest like growth or age at maturity. Thus the clarification of their sex-determining systems would be most beneficial from both a scientific and a practical point of view.

Phenotypic sex is generally labile is fish and can be manipulated by a number of different techniques like chromosomal set manipulation using gynogenesis, androgenesis or triploidy (Devlin & Nagahama, 2002; Penman & Piferrer). Hormonal sex reversal is another option. However, direct sex reversal through hormonal treatment is not a commercially acceptable means to alter sex ratios in food fish within the EU (Directive 2003/74/EC).

Sex determination is the genetic or environmental process that establishes the sex of an organism, whereas sex differentiation is the process with which an undifferentiated gonad is transformed into an ovary or a testis (Penman & Piferrer, 2008).

Even though sex determination in mammals and birds is well studied, current knowledge in fish is far from being complete. In general sex-determining systems in gonochoristic species can be classified in the following categories

- Those under genotypic sex determination
- Those under environmental sex determination
- Those under both genotypic and environmental sex determination

Elaborating on the above, Bull (1983) stated that the inheritance of sex is based on major sex factors, minor sex factors and environmental differences, with evolutionary transitions involving changes in the relative strength of these three major effects being possible. Polygenic and genotype-environment interaction have also been observed (Ospina-Alvarev & Piferrer, 2008).

Fundamental similarities in gonad development are evident at the genetic, molecular, cellular, developmental, and physiological levels in all vertebrates, with genes downstream the major sex-determining locus appearing to play a conserved role (Rhen & Schroeder, 2010). In mammals, sex determination involves a differentiated XX/XY system in which a gene on the Y chromosome (*Sry*) is the major sex-determining gene (Ferguson-Smith, 2006). Genes involved in testis development are often dosage-dependent, where changes in expression levels, in addition to mutations, lead to sex reversal (i.e. XY females and XX males). Additional genes, not located on the Y chromosome, while functional in both sexes could exert a specific effect on the development of the testis (Mittwoch, 2006).

In gonochoristic species with differentiated sex chromosomes (XX/XY, ZW/ZZ), variation in sex chromosome morphology is mainly due to differences in heterochromatin levels, chromosomal fission and rearrangements caused by Robertsonian or tandem fusions. Sex-determining systems evolved also due to gene duplication, retrotransposition, or by insertion/deletion events (Ezaz et al. 2006).

In most fish species, the sex chromosomes are still in early stages of differentiation compared to mammals, and do not show distinct differences in length or gene content. Both XX/XY and ZW/ZZ systems have evolved repeatedly in various groups of fish (Devlin & Nagahama, 2002). Differentiated sex chromosomes, which are visible in mitotic karyotypes are rare even among those species for which there is genetic evidence for monofactorial sex determination (XX/XY or ZW/ZZ) (Harvey et al., 2003a). Morphologically differentiated sex chromosomes have been identified only in about 10% (176 species out of 1700 species) of fish studied karyologically (Ezaz et al., 2006). Additional autosomal loci also contribute to sex determination in many species (Lee et al. 2004).

Variation in sex-determination can exist even between different populations of the same species as is the case with the platyfish *Xiphophorus maculates* where females can be WX, WY or XX and males XY or YY in natural populations (Froschauer et al. 2002). Sex determination mechanism variation could in some cases be due to response to sexual conflicts as has been the case for cichlids from lake Malawi (Roberts et al., 2009).

In general heteromorphic sex chromosomes can appear as different in size or shape when viewed under the microscope, or by using techniques such as banding. By contrast in species with homomorphic sex chromosomes no differences can be observed in terms of size, shape, banding. In this case sex chromosomes can be deduced by studying sex-specific or sex-linked genetic markers or by special crosses involving hormonally sex-reversed individuals (Penman & Piferrer, 2008).

Differentiated sex chromosomal pairs display certain similar characteristics, with one chromosome being largely heterochromatic with suppressed recombination and reduced gene content (Ferguson-Smith, 2006). Suppression of recombination in the sex-determining region is considered as one of the early steps toward the differentiation and has been reported for several fish species. In general suppression of recombination allows the emerging heterogametic sex chromosome to sustain its identity by keeping together genes with functions advantageous for one sex and avoids their transfer, where they might have negative effects on the opposite sex (Schartl, 2004; Volff et al., 2007). It is predicted that the DNA sequence of these sex chromosomes will diverge and the non-recombining region will increase (Harvey et al., 2003a).

Once recombination is suppressed, intra-chromosomal inversions and deletions and mobile sequence elements tend to accumulate in the non-recombining region of the Y or W chromosome. These physical changes to the sex chromosome result in heteromorphy seen in metaphase chromosomal spreads, although it is not possible to state *a priori* whether the hemizygous sex chromosome will be the larger or smaller

chromosome of a heteromorphic pair (Ross et al., 2009). Generally, sex chromosomes in fish, and particularly Y-chromosomes, are rich in transposable elements and other types of repeated elements. Differential accumulation of transposable elements, between the two sexes might play a role in the differentiation of sex chromosomes. Interestingly, the sex-determining region of numerous fish is (sub)telomeric, with evidence from many organisms indicating that chromosome ends are regions of accelerated evolution (Volff et al., 2007).

Of great interest is the fact that spontaneous or hormonally sex-reversed fish are generally fully fertile. The YY genotype is not compatible with life in mammals, however YY and WW genotypes are viable in most fish species. This points to the fact that the gene content of the Y and W chromosome is, for most fish species, very similar to that of their X and Z counterparts (Piferrer & Guiguen 2008; Volff et al., 2007).

Different theories have tried to explain the evolution of sex determining systems. Male and female heterogamety resulting from XX/XY and ZW/ZZ sex-determining systems have been thought to evolve independently from a common ancestor with environmental sex determination, or through an intermediate stage in which sex is determined by environmental factors. However, there are examples, which support the alternative theory that direct ZW to XY transitions have occurred at least in fish and amphibians, and probably in reptiles (Ezaz et al., 2006).

1.1.2 Master sex-determining genes – Sex determination systems in model fish organisms

Master sex-determining genes have been identified mainly in the last decade for a number of teleosts. Usually in the case where the major sex-determining gene resides on the heterogametic member of the sex chromosomal pair, then its presence/absence will generally determine phenotypic sex. However, if the major sex-determining gene is located in homogametic sex chromosomes then phenotypic sex will be determined by a dosage-based mechanism (Liew & Orban, 2013).

Sex-determination in the medaka (*Oryzias latipes*) has been studied in detail. The male (XY) is the heterogametic sex and linkage studies revealed a sex-determining region of reduced recombination in the Y chromosome. The X and Y-chromosomes are homomorphic and morphologically indistinguishable. However, medaka sex chromosomes can be visualized by fluorescent in situ hybridization (FISH) using specific molecular probes (Volff et al., 2007). The medaka is the first fish for which a master sex-determining gene (*dmrt1bY* or *dmy*) was identified. This gene is a Y-specific copy of the autosomal *dmrt1* gene located on LG (Linkage Group) 9. Females have only two autosomal copies of *dmrt1*, while males possess also a Y-linked copy. Mutations in *dmrt1bY* resulted in XY (sex-reversed) females (Matsuda et al., 2002). Molecular analysis suggests that the Y-linked gene in medaka is a recent duplication (Fergusson- Smith, 2006). However, this gene is not a universal sex-determining gene for fish because it is not found on the relevant sex chromosomes of closely related species such as O. *luzonensis* (Sekido & Lovell-Badge, 2008). Interestingly Nanda et al. (2003) showed that other unidentified

autosomal loci are involved as well in sex-determination of *Oryzias latipes*, since a significant number of males were found to be lacking the *dmrt1bY*.

Dmy has been detected only in *O. latipes* and *O. curvinotus* out of the 20 *Oryzias* species. *Gsdf*, a gene downstream of *Dmy* in the sex-determining pathway, was shown to have replaced *Dmy* as the master sex-determining gene in *O. luzonensis*. As opposed to previous belief that the master sex determining genes, including *Dmy*, encode transcription factors *Gsdf* encodes a secretory protein belonging to TGF-β superfamily (Myosho et al., 2012).

A duplicated copy of the anti-Mullerian hormone (amh) gene, a TGF-β superfamily member like *Gsdf* above, and a well-characterized hormone in mammals, plays the key role in primary sex determination in *O. hatcherii* (Hattori et al 2012). *Amhr2* (anti Mullerian hormone receptor II), another member of TGF-β protein family, showed complete association with phenotypic sex in tiger pufferfish, *Takifugu rubripes* (Kamiya et al., 2012).

Genome-wide linkage mapping in three-spined stickleback identified a chromosomal region in LG 19, as being responsible for sex determination (Peichel et al 2004). In the study of Kitano et al. (2009) it was also concluded that LG 19 is the sex chromosome. However, several LG9 markers co-segregated with LG19 markers previously found to be tightly linked to the sex determination locus in a region of reduced recombination and rearrangements on the Y chromosome. This association was observed when male meioses, but not female meioses were analysed, indicating

that one copy of LG9 might be fused to one copy of LG19 forming a neo-Y chromosome in Japan Sea sticklebacks. In the nine-spined stickleback (*Pungitius pungitius*) a region in LG12 has been shown to be associated with sex determination. However, in brook stickleback (*Culaea inconstans*) and in four-spined stickleback (*Apeltes quadracus*) neither LG12 nor LG19 were linked to sex (Ross et al., 2009).

The zebrafish (*Danio rerio*), one of the major non-mouse models for the study of vertebrate development, possesses a complicated sex-determining system. In the study of Anderson et al. (2012) it was indicated that zebrafish has a polygenic sex determination system, with probable interactions with environmental factors. A region in chromosome 4 showed characteristics of a sex chromosome, like suppression of recombination and high heterochromatin levels. Finally almost no information about sex determination exists in the spotted green pufferfish *Tetraodon nigroviridis* (Volff et al., 2007). Yano et al. (2012a) discovered that *sdY*, a gene previously associated with immune related functions in other organisms, is the master sex-determining gene in rainbow trout (*Oncorhynchus mykiss*). The *sdY* gene is highly conserved as the major sex-determining gene in most salmonids (Yano et al., 2012b).

1.1.3 Temperature sex determination in fish

Sex is regarded to be determined once an embryo's fate is set under normal conditions of development whereas sex differentiation is the process that follows (Georges et al., 2010). The environment may influence gene expression via the

neuroendocrine system, altering hormonal production, or due to direct environment alterations of hormone production (Gilbert, 2005). Environmental effects on sex ratios may vary even among different strains of the same species (Mylonas et al., 2005). The most important environmental factors involved in sex determination are temperature, pH, photoperiod, and salinity (Devlin & Nagahama, 2002). Interestingly in the study of Mankiewicz et al. (2013) background colour was shown to have a significant effect in the sex-determining system of southern flounder (*Paralichthys lethostigma*).

Temperature has been the factor studied in most detail in fish as a result of observations in reptiles where temperature-dependent sex-determination (TSD) is quite common (Pieau, 1996). In gonochoristic vertebrates, sex-determining mechanisms can broadly be classified as genotypic (GSD) or temperature-dependent (TSD) (Ospina-Alvarez & Piferrer, 2008).

TSD is a case of ESD (Environmental Sex Determination) where temperature during sensitive periods of early development can irreversibly determine phenotypic sex. In fish, temperature effects have been described in 8 families of jawed fish plus one Agnathan species (Devlin & Nagahama, 2002) and have been reported to be present in several species of cultured fish (Baroiller & D'Cotta, 2001). Probably the most distinctive chronological difference in development between GSD and TSD is in the timing of commitment to each sex (Valenzuela, 2008). In species with TSD, there are no consistent genetic differences between sexes. In temperature-dependent sex determination (TSD), the thermal conditions during a critical phase in early life can induce the formation of functional ovaries or testes (Hattori et al., 2007).

Fish with TSD have been grouped according to three groups of sex ratio response to environmental temperature 1) more males at high temperatures 2) more males at low temperature 3) more males at extreme (high and low) temperatures. The first group is the most common one with the number of fish species assigned to the three groups above currently amounting to 53-55, 2-4 and 2 respectively (Ospina-Alvarez & Piferrer, 2008). The increase in males at high temperature could be the result of sex-reversal of females as a consequence of the inhibition of aromatase, the enzyme that converts androgens to estrogens. However, it has been pointed out that the sex ratio shifts above might be the consequence of thermal effects on GSD (GSD + TE) rather than the presence of TSD (Ospina-Alvarez & Piferrer, 2008).

According to Valenzuela et al. (2003) and Conover (2004) any given species in order to be considered as having a TSD instead of GSD +TE, should fulfil both of the following two conditions: 1) not having sex chromosomes, and 2) have sex-ratio response to temperature within the range of temperatures that is expected to encounter in the wild. Analysis of both field and laboratory data for the 59 species of fish, where TSD has been explicitly or implicitly assumed, showed that TSD in fish is less frequent than previously thought (Ospina-Alvarez & Piferrer, 2008).

1.2 Introducing the farmed species used in the current study

A wide variety of fish species in aquaculture exhibit sexual dimorphism in a range of traits of interest like growth or age at maturity. Thus the clarification of their sex-determining systems would be most beneficial from both a scientific and a practical point of view. The present study focuses on three farmed species of the above category. Atlantic halibut (*Hippoglossus hippoglossus*), Nile tilapia (*Oreochromis niloticus*) and European sea bass (*Dicentrarchus labrax*). The species above were chosen with the aim of covering species with varying level of complexity of the underlying sex determining system.

1.2.1 Atlantic halibut (*Hippoglossus hippoglossus*)

1.2.1.1 Background information

Flatfish comprise a unique group of teleosts due to a unique developmental process known as metamorphosis. Flatfish show a range of sex-determining mechanisms, including XX/XY and ZW/ZZ, with significant effects of environmental factors, principally temperature, in some species (Babiak et al., 2012). The aquaculture production of turbot (*Psetta maxima*) is the highest among flatfishes, whereas that of Atlantic halibut (*Hippoglossus hippoglossus*) is now successfully underway (Cerda et al., 2010). Atlantic halibut has the highest growth rate of any flatfish making it a promising candidate for aquaculture (Hendry et al., 2002). It has been a high-value species for cold-water marine aquaculture for several decades in Northern Europe

and America, although production has been limited by a series of bottlenecks. Sexual dimorphism in growth, with males maturing earlier and growing significantly slower than females, reduces productivity and profitability of the sector. Females are expected to reach market size at around 36 months while males would need at least an extra year making the production of all female stocks particularly appealing for the aquaculture industry (Babiak et al., 2012; Bjornsson, 1995).

1.2.1.2 Sex determination of Atlantic halibut

Gynogenetic studies in *H. hippoglossus* indicated that this species may possess an XX/XY sex-determining system. Gonadal sex-differentiation occurs at around 38 mm fork length, which corresponds to the metamorphic and post-metamorphic stages (Hendry et al., 2002). Previous studies have shown that it is possible to significantly skew the physiological gender of a population using hormonal treatments (Hendry et al., 2003; Babiak et al., 2012). However, such methods are not acceptable means to alter sex-ratios in food fish within the EU (Directorate General for Health and Consumers, 2003) thus indirect sex reversal is required whereby masculinised genotypic females (neomales) are crossed with normal females to create all-female progeny.

Currently the main technique for such verification is progeny testing of treated animals which is time consuming and costly, a process which takes at least 4-5 years in halibut due to the generation interval of over 3 years. Direct genetic sexing would be preferable using non-lethal and cheap genotyping techniques, at least in the simple cases of male or female heterogamety. Rearing temperature does not seem have an effect on phenotypic sex in this species (Hughes et al., 2008). No information exists concerning the locations of sex-determining regions in Atlantic halibut.

1.2.1.3 Atlantic halibut genomic resources

Hippoglossus hippoglossus has limited genomic resources available. The most important resource is the available genetic map, which contains 258 microsatellites and 346 AFLPs grouped in 24 linkage groups corresponding to its karyotype (Reid et al., 2007). More than 20,000 ESTs are available from databases like NCBI.

1.2.2 Nile tilapia (*Oreochromis niloticus*)

1.2.2.1 Background information

Tilapias belong to the family Cichlidae and are abundant in tropical and subtropical regions. The fact that tilapias are sturdy fish makes them well suited to aquaculture under a wide range of conditions. Some of the most important attributes of tilapias that makes them ideal candidates for aquaculture include fast growth, tolerance to a wide range of environmental conditions, resistance to stress and disease, ability to reproduce in captivity, short generation time, feeding at low trophic levels and acceptance of artificial feeds immediately after yolk-sac absorption (El-Sayed, 2003).

Tilapias rank second in world aquaculture production after carps (FAO 2012). The most important in terms of production are the Nile tilapia (*Oreochromis niloticus*) and the blue tilapia (*Oreochromis aureus*). In 2010 the world aquaculture production of tilapias exceeded 2.8 million metric tones. The main tilapia producers worldwide are China, Egypt, Philippines, Indonesia, Thailand, Taiwan and Brazil (FAO 2012). The edible part of the carcass of male and female tilapia consists of ~53% and 51% respectively. Fillet yield is about 40% of the whole fish body weight (Dikel & Celik, 1998).

In the tropics tilapias can breed all year round, while in the subtropics the breeding season is limited by water temperature and may range from 5 to 7 months. Females can breed every 3-4 weeks during the breeding season after attaining their sexual

maturity while still being smaller than marketable size (Cnaani & Hulata, 2008a). Intensive commercial production generally requires all-male stocks, not only because males grow faster but also to avoid uncontrolled reproduction before harvest. Mono-sex populations are currently usually produced using steroid hormone treatment. This is undesirable for a number of reasons with the most serious one being legislation issues in the European Union concerning usage of hormones in food. Another method, which avoids the drawbacks of the above, involves the production of the so-called "genetically male tilapia" (GMT) that was coined by Mair et al. (1997). Genetically male tilapia is produced from crosses between YY males and YY neofemales that have been identified after a series of crosses and progeny testing. Crossing YY males with ordinary females for mass production is expected to produce all female progeny (Fig 1-1).

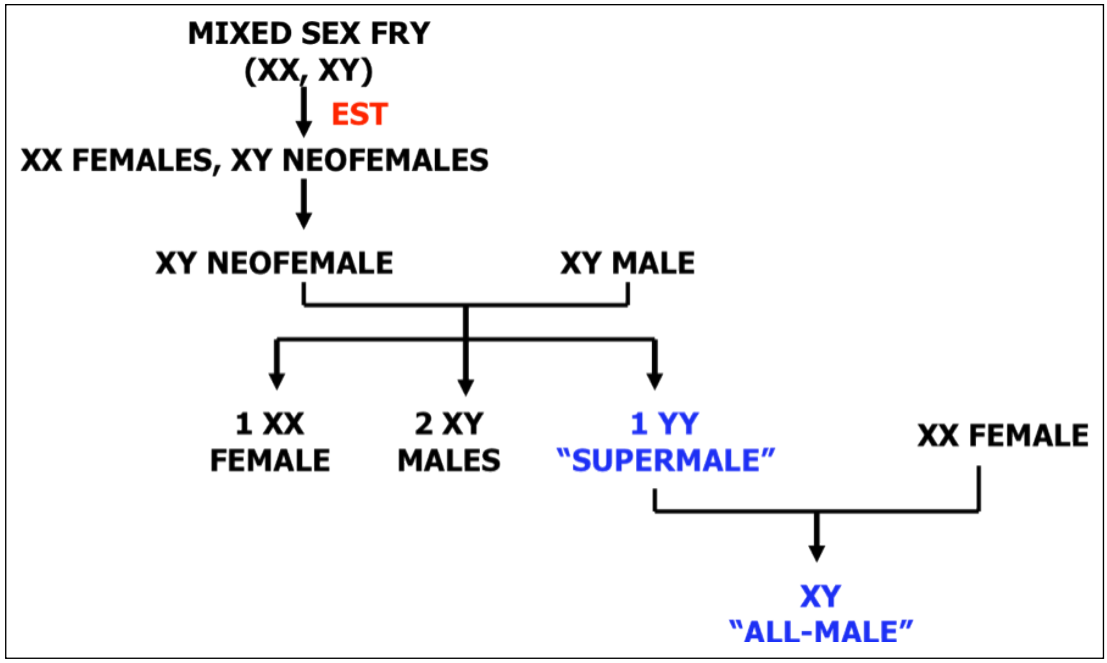

Figure 1-1. Production of genetically male tilapia. EST: 17β-estradiol or diethylstilbestrol. Modified from Mayer et al 1997.

1.2.2.2 Sex-determining system in tilapias

No significant morphological differences in any chromosomal pair have been observed so far in tilapias. Research suggests the existence of two sex-determining systems among *Oreochromis* species with some species possessing a XX/XY system (*O. mossambicus*, *O. niloticus*) whereas in others there is a ZW/ZZ system (*O. aureus*, *O. macrochir*, *O. urolepis hornorum, O. tanganicae*). The primary support for these hypotheses comes from breeding animals that were sex-reversed by hormone treatment, (Shirak et al., 2006). Male heterogametic systems were suggested in *O. mossambicus* and *O. niloticus*, when crosses of sex-reversed (XX) males with normal (XX) females produced only females. In *O. aureus* mating between sex-reversed (ZZ) females and normal (ZZ) males usually results in 100% male offspring (Cnaani et al., 2008b).

Hybridization between some species of tilapias as *O.niloticus* and the blue tilapia, *O.aureus*, results in the production of near all-male offspring. However, the potential of tilapia hybrids for culture presents difficulties mainly related to the sourcing and management of the pure species. Without careful broodstock management females start to appear in what was previously an all-male hybrid. Alternatives for production of all-male tilapia involve the hormonal sex inversion and progeny testing to identify YY males (Cnaani & Hulata, 2008a). YY male Nile tilapias are used as broodstock in YY x XX crosses to produce genetically male tilapia for aquaculture (Fig 1-1, Mair et al., 1997). In general YY male genotypes of Nile tilapia can be as viable and

as fertile as normal XY males, and sire progeny that are nearly 100% males (Mair et al., 1997).

Selection for sex-ratio (high percentage of males) has been applied within the male and female broodstock lines, and the GMT technique has been tested in different strains of *O. niloticus* (Tuan et al., 1998, 1999). A YY male line was further selected for growth rate using intensive within-family selection (Abucay & Mair, 2007). However, YY male/XX female crosses often produce less than 100% males, and empirically it has been shown that having more than about 5% females will lead to the production of unwanted fry during grow-out in many aquaculture systems. Both genetic and environmental factors (principally temperature) can influence sex-ratios away from that expected from the main XX/XY system in this species (Penman & Piferrer, 2008).

Current evidence suggests that O. *niloticus* possess an XY/XX male heterogametic system controlled by genetic, environmental factors (with temperature being the most important) and probably by their interaction. Using the genetic map of Lee et al. 2005, that was derived from an interspecific cross between Nile and blue tilapia, different studies presented evidence of different sex-determining regions. Lee et al. (2003) mapped a sex-determining Quantitative Trait Locus (QTL) in in an 10 cM interval in Linkage Group 1 (LG1) using microsatellite markers. However, the association persisted in only two out of the three crosses studied. Lee et al. (2011) identified three sex-linked AFLP markers also in LG 1, but they were outside the previous estimated interval for the location of the major sex-determining QTL.

Karayucel et al. (2004) showed evidence for linkage between the red body colour gene and XX males in Nile tilapia. The red body colour gene has been mapped in LG3 (Lee et al., 2005), the same LG as the ZW/ZZ locus in blue tilapia (Lee et al., 2004).

On the other hand in the studies of Eshel et al. (2010; 2012), even though sex associated markers were found on LG 1, markers on LG 23 were the ones that showed the highest association with phenotypic sex. Eshel et al. (2010) used a full-sib family (156 individuals), which was genotyped for microsatellites across different linkage groups (1, 2, 3, 6, 23). The sex-determining QTL was mapped in a region spanning 16-21 cM in LG 23. Eshel et al. (2012) used additional markers to fine map the previous QTL. The QTL was also mapped in terms of physical distance on the O. *niloticus* draft genome, in a region of 1.5 Mb in scaffold 101. In this study they only genotyped a single family derived from a cross between a normal male and a sex reversed female and presented no information on whether the above QTL on LG 23 persists in other crosses.

Lee et al. (2004) analyzed a single cross of blue tilapia (WZ/ZZ) and concluded that there was a primary sex-determining region in LG3 and a secondary sex-determining region in LG1. One third of the individuals with the ZZ/ZZ genotype under this model were female, compared to none of the fish with the ZZ/XY genotype. Cnaani et al. (2004) identified microsatellites associated with sex in LGs 1, 3, and 23 in an interspecific tilapia hybrid derived from a cross between Mozambique tilapia (*O. mossambicus*) and blue tilapia (*O. aureus*). In *O. tanganicae* microsatellites from LG

3 were associated with phenotypic sex (Cnaani & Kocher, 2008). Liu et al. (2013) mapped the sex-determining region of *Oreochromis mossambicus* and of a red tilapia strain (hybrid between Nile and Mozambique tilapia) in LG1 and LG23 respectively.

Phylogenetic analysis suggests that there have been repeated transitions between LG3 and LG1 as sex chromosomes in tilapia, and it has been suggested that the sex-determining locus in LG3 (chromosome pair 1) is the ancestral sex-determining gene in this group (Ocalewicz et al., 2009). FISH mapping showed that LG3 corresponds to the largest pair of chromosomes and LG1 to a small pair, in contrast to earlier cytogenetic (meiotic chromosome pairing) evidence suggesting that the *O. niloticus* sex-determining locus was located in the largest pair of chromosomes (Cnaani et al., 2008b). LG3 has more of the characteristics of a sex chromosome pair (recombination suppression as well as the accumulation of repetitive elements) than the chromosome pair carrying the proposed newer sex-determination locus in LG1 (Ocalewicz et al., 2009).

1.2.2.3 Temperature effect on tilapia sex determination

Sex reversal has been used both in the elucidation of sex-determining mechanisms and in production of monosex populations for aquaculture. Apart from hormonal treatment environmental factors can also induce sex-reversal. Although a number of environmental factors have been implicated in influencing sex-ratio such as pH, stocking density, pollution and social interactions, in tilapias temperature seems to be the most important and best studied one (Beardmore et al., 2001).

Despite evidence for a genetic basis for determining sex in tilapia it is clear that other factors are also acting on sex, with a strong effect of temperature on sex-differentiation having been demonstrated in various tilapia species (Baroiller et al., 2009). Studies in Nile tilapia have shown that rearing at high temperatures (36^0C) could efficiently masculinise some progenies if the treatment started around 10 days post fertilization (dpf) and was applied for at least 10 days with longer periods being as effective (Baroiller et al., 1995; Baras et al., 2001; Baroiller & D'cotta, 2001; Tessema et al., 2006; Wessels & Horstgen-Schwark, 2007). However, tilapias did not pass the required criteria to be considered TSD species in the study of Ospina-Alvarev & Piferrer (2008) where they were considered as being a prime example of GSD + TE species, but not TSD species.

The window for temperature sensitivity coincides with the gonad sensitivity towards other external factors, notably hormones. Like temperature, hormonal treatments or the use of aromatase inhibitors during sex-differentiation can override the genetic sex-determination, inversing sex and producing functional phenotypes (Baroiller et al., 2009).

The temperature sensitivity of Nile tilapia during sex-differentiation is not seen in all progenies. Some male or female breeders produce progenies displaying a high sensitivity to temperature giving a high proportion of males in their sex ratio, while others seemed to be insensitive to elevated temperatures (Baroiller et al., 2009). A study from Baroiller & D'Cotta (2001) demonstrated that the percentage of males was very different at the inter-individual level, indicating that there was an important

parental effect. Temperature induced sex-determination was also confirmed when natural populations of *O. niloticus* were studied by Tessema et al. (2006) and Bezault et al. (2007). Wessels & Horstgen-Schwark (2007) provided evidence that a surplus of males in temperature treated groups could be selected for, as a quantitative trait. A heritability of 0.69 was obtained through a two-generation selection experiment for high temperature sensitivity

As already mentioned the major genetic locus in Nile tilapia is present on LG1 but other loci (LG3, LG23) are probably acting simultaneously to determine sex and despite the fact that the fate of the gonad is determined genetically, temperature can override it. The really critical period of gonad differentiation in tilapia has been established to be from 9 to 15 dpf (Wessels & Horstgen-Schwark, 2007; Ijiri et al., 2008). Temperature or hormonal treatments have to be applied from this period onwards to be efficient. This is just before the appearance of the very first sex-specific difference, an active mitosis in the ovary (D'Cotta et al., 2001; Ijiri et al., 2008). Interestingly there is no explanation yet as to why at extreme temperatures an XX individual would have a better fitness as a male rather than as a female (Baroiller et al., 2009).

1.2.2.4 Role of known genes involved in sex determination

Several genes known for their role in the sex-determining pathways in other species have been mapped in tilapia. Five of these genes were found to be located close to other sex-determining loci (Cnaani et al., 2007,2008b):

- Aromatase Enzyme *Cyp19a1* on LG1
- *Wt1b* on LG1
- *Amh* on LG23
- *Dmrta2* on LG23
- *Sox14* on LG23

However, none of the above seems to be the major sex-determining gene(s) in tilapia (Cnaani & Hulata, 2008a).

Cyp19 is the enzyme that catalyzes the irreversible conversion of androgens into estrogens. Since all estrogens derive from androgens, and *cyp19* is at the end of the steroidogenic pathway it determines the balance between these two types of steroids (Piferrer & Guiguen, 2008). Ijiri et al. (2008) found higher levels of *cyp19a* in future ovaries of Nile tilapia as early as 9 dpf (day post fertilization) and increased until 19 dpf. Immunohistochemical studies carried out in tilapia gonads showed that, in females, all the major steroidogenic enzymes, including *cyp19a*, are present before the onset of sex differentiation (Piferrer & Guiguen, 2008). Likewise, in species where temperature can affect sex differentiation, *cyp19a* is present at feminizing temperatures but suppressed at masculinising temperatures. Temperature applied during the sex differentiation period in tilapia XX offspring induced a down-

regulation of *cyp19a* (Baroiller et al., 2009). The physical mapping of the aromatase gene does not place it near the sex-determining region of *O. niloticus*. However, does not exclude the possibility that the above gene being one of the autosomal genes that can affect sex ratios in this species (Harvey et al., 2002b, 2003ab).

Wt1b encodes a zinc-finger DNA-binding protein involved in testis development and is shown to up-regulate SRY expression by binding to DNA (Miyamoto et al., 2008). *Wt1b*, like *cyp19a* is located on LG1 of Nile tilapia in close proximity to the major sex-determination locus, but linkage analysis studies excluded Wt1b as the sex determinant gene (Baroiller et al., 2009; Lee & Kocher, 2007).

Amh enzyme is responsible for the regression of the Mullerian ducts in males. However, *Amh* was found in fish, which do not have Mullerian ducts, and its role is still unknown. *Amh* in tilapia increases in XY male gonads before that of either *Sox9a* or *Sox9b*, similar to reports on chicken and on reptiles with TSD. Since Amh is upregulated in supporting cells it probably has a role in testicular differentiation. It is interesting that *Amh* has been mapped to LG23 where two QTLs for sex have been located (Shirak et al., 2006; Baroiller et al., 2009).

Dmrt, although not thoroughly analyzed yet in tilapia, is believed to play a critical role in male differentiation (Ijiri et al., 2008). This gene is closely related to the male determinant *DMY/Dmrt1bY* of medaka and has a male-specific expression at early stages in chicken and turtles during testis development but does not seem to play a role in mammalian testis determination (Baroiller et al., 2009).

Sry is considered the master key regulator of sex determination in mammals, but this gene has not been found in most other vertebrates. *Sry*-related (*sox*) genes have recently been studied in various fish species and encode a family of transcription factors that are involved in sex determination and other developmental processes in vertebrates (Shirak et al., 2006). In tilapia *sox* genes expression levels in XX and XY gonads were similar from 9 to 29 dpf in the study of Ijiri et al. (2008) becoming stronger thereafter in XY males. Cnaani et al. (2007) mapped *sox14* in LG23 in four different tilapia species and although it has been excluded as the major sex-determining locus it was within a region previously believed to have an effect in sex-determination (Shirak et al., 2002; 2006).

1.2.2.5 Nile tilapia genomic resources

Oreochromis niloticus has rich genomic resources available. Most importantly its genome is available (Bouillabase.org). The last version (Orenil1.0) grouped sequenced scaffolds in 22 chromosomes, corresponding to its karyotype. The size of the sequenced genome is 927.74 Mb in which 23,080 genes have been identified. A genetic map containing 525 microsatellites grouped on 24 linkage groups (Lee et al., 2005) and a radiation hybrid map (Guyon et al., 2012) containing 1296 markers (genes, microsatellites and SNPs) grouped in 81 RH (Radiation Hybrid) groups are also available. Finally more than 120,000 ESTs are available in databases like NCBI.

1.2.3 European sea bass (*Dicentrarchus labrax*)

1.2.3.1 Background information

European sea bass has a high commercial value, both from captures from wild stocks and from aquaculture production in the last 30 years. Greece and Turkey are the largest producers, with an annual production of ~43,000 t each (FEAP 2012). It has a production cycle of 18-24 months at the end of which it is usually marketed at around 350-400 gr. The sea bass is a eurythermal fish with high tolerance to salinity changes. It is a gonochoristic species spawning in the Mediterranean Sea between December to March and in the Atlantic Ocean in June. First sexual maturity occurs during the second year for males (23-30 cm) and during the third year for females (31-40 cm). Female gonads complete maturation at the same time, releasing all eggs together in short time. A mature female may produce between a quarter and half a million eggs per kg of body weight (Volckaert et al., 2008).

Fry remain sexually undifferentiated for a long period, with the differentiation occurring between 128 days to 250 days post fertilization. Sex differentiation proceeds in a caudo-cranial fashion usually starting at 200 days post-hatching (dph), with females differentiating first. Records of sea bass sex ratio in wild populations though scarce, provide evidence for a balanced sex ratio (Vandeputte et al., 2007). Under culture conditions a high percentage of males (70-90%) is usually observed. Males generally grow more slowly than females, with studies showing that males

could be even 40% lower in body weight at harvest time (Gorshkov et al., 2004a; Zanuy et al., 2001). Saillant et al. (2003b) showed that females start growing faster before morphological sex-differentiation occurs. In the study of Vandeputte et al. (2007) 1-year old females were heavier by approximately 41% and 2-year old females by approximately 20% than males of the same age group. In addition precocious males grow even slower and can weigh up to 18% less than non-precocious males by the time of marketing (Navarro-Martin et al., 2009a).

Complete masculinisation is possible by hormone treatments in early ontogenesis, using the synthetic androgen 17α-methyldıhydrotestosterone or through 17α-methyltestosterone (Blazquez et al., 2001; Chatain et al., 1999). An increase percentage of females can also be induced through similar treatments with either the natural estrogen 17β-estradiol or with the synthetic estrogen 17α-ethynylestradiol(Gorshkov et al., 2004b). Developing methods to manipulate the sex of European sea bass would become possible only after a thorough understanding of its sex-determination and differentiation systems (Papadaki et al., 2005).

1.2.3.2 Sex determination in European sea bass

The karyotype consists of 24 subtelocentric-acrocentric chromosome pairs with no recognizable heterochromosomes. Meiotic gynogenesis in sea bass indicated that the mechanism of sex determination does not correspond to a simple monofactorial system with female homogamety (Peruzzi et al., 2004). Also the sex ratio of the

offspring from hormonally masculinised females is not female biased and would rule out both XX/XY (female homogamety) and ZW/ZZ (male homogamety) systems (Blazquez et al., 2001). However, the sex-ratio of progeny from gynogenetic females was skewed in favour of females (Francescon et al., 2005).

Sexual dimorphism in length and weight has been observed in sea bass even before sexual maturity (Saillant et al., 2003b). This dimorphism could be explained by the fact that male sea bass reach sexual maturation one year earlier than females, thus expending relatively more energy on gonad growth (Koumoundouros et al., 2002). Strong parental additive influences with genotype-temperature interactions can modulate the sex-ratio in sea bass. The proportion of females resulting from individual crossings may vary from 1 to 70% (Piferrer et al., 2005; Sailant et al., 2002). Using successive size gradings Papadaki et al. (2005) were able to produce female dominant and male dominant populations.

Current information suggests that the sea bass sex determination system is polyfactorial, depending on both genetic factors and environmental (temperature considered as the most important one) effects (Navarro-Martin et al., 2009ab; Piferrer et al., 2005), with no master sex-determining gene being present (Deloffre et al., 2009). In the study of Saillant et al. (2003a) it was hypothesized that apart from temperature, salinity level could indirectly influence the sex mechanism through a complex environmental effect. In the study of Vandeputte et al. (2007) sex was treated as a polygenic trait with both sire and dam having an effect on sex-ratio. Its heritability was estimated at 0.62 (s.e. 0.12). In the study of Chatziplis et al. (2007)

the heritability of phenotypic sex was estimated to be only 0.12. However, since all the families used to calculate the heritability had a common dam, this heritability estimate is less reliable. Traits with high estimated heritability, such as body weight and sex, can be easily selected for in order to achieve strong genetic gains through classical breeding designs. Carefully designed breeding programs with pedigree information and Best Linear Unbiased Predictor (BLUP) estimates of breeding values can lead to a 10% genetic gain per generation without compromising the effective population size and level of inbreeding (Volckaert et al., 2008). No genetic markers associated with phenotypic sex have been found until now in sea bass. However, in the study of Blazquez et al. (2008) it was shown that *cyp19a* expression levels could be used to discriminate phenotypic females.

1.2.3.3 Temperature effect on the sea bass sex determination

Temperature clearly influences sea bass sex-ratio, with high temperatures favouring the development of males (Volckaert et al., 2008; Piferrer et al., 2005). The influence of temperature on sex differentiation seems to be more complicated than other models in reptiles or fish. In contrast to other species P450 aromatase activity was not influenced at masculinising temperatures (Socorro et al., 2007). However, in the study of Navarro-Martin et al. (2011), where different activity was observed between groups in high and low temperature, differences in the methylation level, of the promoter of *cyp19a* were considered to be the underlying mechanism behind the male skewed sex-ratios observed after rearing in high temperatures.

In the study of Koumoundouros et al. (2002) sex-ratio was correlated with the growth rate of the fish up to the end of the TSD-sensitive period with the larger fish presenting a significantly higher ($P < 0.01$) female incidence than the smaller fish in the tested thermal regimes. High temperatures (~ 21^0C), which are typically used during the larval and early juvenile stages are thought to cause sex-reversal of genotypic females. Mylonas et al. (2005) concluded that decreasing the water temperature from 21^0C to 15^0C increases the female proportion of the population, while further decreasing the rearing temperature below 15^0C does not cause any further increase. In the study of Navarro-Martin et al. (2009a) it was shown that high temperatures masculinize on average over half of the females. In this study it was proposed that rearing at 17^0C from fertilization for 850-900 degree days, would result in an increase of biomass of over 10% at harvest. Combined with genetic selection for higher female ratios an even larger increase of biomass could be achieved (Vandeputte et al., 2007).

1.2.3.4 European sea bass genomic resources

The average genome size of sea bass, estimated by flow cytometry is approximately 760 Mbp. Using whole genome shotgun sequencing Kuhl et al. (2010) produced an approximate two-fold coverage of the sea bass genome, while a newer version is about to become available in the near future (R. Reinhardt, 2014, personal communication). A genetic map is also available with its last version consisting of 190 microsatellites and 176 AFLPs (Chistiakov et al., 2008).Additionaly a radiation

hybrid map is also available (Guyon et al., 2010). More than 86,000 ESTs are available from databases like NCBI.

1.3 Application of genetic markers in the analysis of the sex determination system of fish

Analyzing phenotypic sex-ratios of fish produced by methods such as hybridization and chromosome set manipulations (gynogenesis, androgenesis), and progeny testing of sex-reversed individuals has allowed identification of simple genetic sex-determining systems (Penman & Piferrer, 2008). Currently the main techniques for producing mono-sex stocks in fish require progeny testing of hormonally sex-reversed animals. The above is time consuming and costly, especially in species with delayed maturity.

The advent of DNA analysis techniques has facilitated the search for sex-linked and sex-specific sequences. In general there is still a lack of molecular markers associated with phenotypic sex for the majority of species, that if available could be used to genotype sex in a number of aquaculture and model species, facilitating the production of mono-sex stocks (Piferrer & Guiguen, 2008).

Genetic approaches used include:

- Comparison of male and female DNA profiling/fingerprinting
- Subtractive techniques to look directly for sequences which differ between the male and female genomes

- Candidate genes, where genes or sequences that are sex-determining or sex-linked in one species are analyzed in the target species
- Linkage mapping
- Molecular-cytogenetic (Penman & Piferrer, 2008).

With genetic markers, it became possible to observe genetic variation in the entire genome. Popular genetic markers in aquaculture research include allozymes, mitochondrial DNA, Restriction Fragment Length Polymorphism (RFLP), Random Amplified Polymorphic DNA (RAPD), Amplified Fragment Length Polymorphism (AFLP), microsatellites and Single Nucleotide Polymorphisms (SNPs). The application of genetic markers has allowed rapid progress in aquaculture investigations of genetic variability and inbreeding, parentage assignment, species and strain identification, and the construction of high-resolution genetic linkage maps for aquaculture species (Liu & Cordes, 2004).

A genetic marker is an identifiable physical sequence on a chromosome or a protein product whose inheritance can be monitored (Hayes & Oivind, 2005). Genetic markers can generally be divided into different categories. Taking into account their mode of inheritance genetic markers are divided into co-dominant and dominant. In the first category, which generally provides the most information per marker, it is possible to distinguish the heterozygous individuals from the homozygous ones for each allele. In this category are the allozymes, RFLPs, microsatellites and SNPs. In dominant markers such as RAPDs and AFLPs it is not possible to identify heterozygotes directly. Genetic markers are considered as Type I if they are associated with genes of known function and Type II if they are associated with

anonymous genomic segments. Allozymes are Type I while the rest can be classified in both categories.

Simple sequence repeats (SSR), also known as microsatellites, are one of the most widely used genetic markers in fish. They are co-dominant, which means that homozygous and heterozygous individuals can be distinguished (Fjalestad et al., 2003). Microsatellites have a number of desirable properties including high levels of polymorphism and consequently high information content and ease of amplification. In fish it is thought that microsattelites occur at a frequency of approximately every 10 kb (Iyengar et al., 2000). On the other hand detection of microsatellites is more difficult than detection of AFLP markers and they are more expensive to genotype. Problems can arise in certain cases due to the appearance of stutter bands, which are believed to be due to polymerase slippage during PCR, making the scoring of the alleles difficult. Genotyping errors can also occur due to preferential amplification of small alleles and due to mutations at primer sites making certain alleles non-amplifiable (null alleles) resulting in false homozygotes (Oosterhout et al., 2004). Microsatellites are generally found in the non-coding part of the genome, and so have no effect on gene expression. One difficulty with microsatellites, which can be a problem in gene-mapping, is that they may not exist in sufficient density in the genome to be very closely linked to genes affecting important traits in aquaculture production (Hayes & Oivind, 2005).

1.4 Next Generation Sequencing (NGS) and Restrition-site Associated DNA-sequencing (RAD-seq)

Next generation sequencing has significantly lowered the cost of whole genome sequencing making possible to apply to a wide range of aquaculture species (Sundquist et al., 2007). Currently sequencing the human genome at 30X coverage costs approximately UK £5,000 (Davey et al., 2011). Surprisingly, the full genome sequencing is not necessarily the optimum way of characterizing the sex-determination system, as numerous genome sequences are now available with no information on such major sex-determining genes even in highly annotated species like the zebrafish. The characterization of sex-specific molecular markers for aquaculture purpose could be a more reasonable and cost effective approach (Piferrer & Guiguen, 2008).

Restriction-site Associated DNA sequencing (RAD-seq) has recently been used with a number of different fish species since the technology became available in 2008. One of the aims of the Baird et al. (2008) study, which first validated the technique in fish, was to detect Quantitative Trait Loci (QTL) in three-spined stickleback (*G. aculeatus)*. A number of different restriction-digest methodologies already existed, using high-throughput sequencers. However, what sets the RAD-seq methodology apart was the fact that it combined control over the fragment size that result from the digestion with deep sequencing across individuals, making the identified SNP reproducible (McCormack et al., 2012). This makes the RAD platform very efficient for constructing genetic maps and QTL studies.

In RAD-seq genomic DNA from multiple individuals is digested with a restriction enzyme of choice and pooled, after ligating adaptors with barcodes for identification of each sample. The resulting restriction fragments are selected for a size most suitable (Illumina sequencing 300-700 bp) and after a subsequent Polymerase Chain Reaction (PCR) step the fragments are sequenced, producing partial but genome-wide coverage at a fraction of the cost of whole-genome sequencing. Illumina sequencing of short fragments involves sequencing one (read 1, single end) or both (reads 1 and 2, paired end) ends of each fragment, typically producing reads 100 bp long (Davey et al., 2012). RAD tags create a reduced representation of the genome, allowing over-sequencing of the nucleotides next to restriction sites and detection of SNPs. By using a restriction enzyme of choice the number of markers can be increased multi-fold. Also the above approach is amenable to genotyping pooled populations for bulk-segregant analysis and also multiplexed genotyping of individuals for fine-scale mapping (Baird et al., 2008; Hohenhole et al., 2010).

Petersen et al. (2012) elaborated on the RAD-seq platform by using a double digest (dd) with two restriction enzymes and eliminating the shearing step (ddRAD-seq). This combined with the usage of a combinatorial multiplex indexing (primers with different coding), allowed several hundred individuals to be pooled in a single sequencing lane. ddRAD-seq has been successfully used in the study of Recknagel et al. (2013) where a linkage map of two cichlids was obtained.

1.5 Aims and outline of the thesis

The general aim of this study was to investigate the genetics of sex determination in farmed fish with important sexual dimorphisms using Next Generation Sequencing. The results of the current study are expected to be of value in future research aiming at the production of mono-sex stocks for aquaculture purposes. The species chosen for the current study show growth dimorphism between the two sexes, while at the same time allowed studying sex-determining systems of varying complexity. Below is a general summary of the topics covered in each of the following chapters:

- Chapter 2 elaborates on the general materials and methods that were followed throughout this thesis.
- Chapter 3 focuses on mapping the major sex-determining region of Atlantic halibut using RAD-seq.
- Chapter 4 focuses on mapping the major sex-determining region of Nile tilapia using RAD-seq.
- Chapter 5 focuses on mapping sex-determining regions in Nile tilapia families with skewed sex ratios using ddRAD-seq.
- Chapter 6 focuses on the study of European sea bass sex-determining system using RAD-seq.
- Chapter 7 summarizes the results of this thesis and discusses the possible applications and directions for further research.

Chapter 2

General Materials and Methods

2 General Materials and Methods

2.1 Nile tilapia (*Oreochromis niloticus*) basic maintenance and handling

2.1.1 General Information

Three different species were used in this study, Nile tilapia, Atlantic halibut and European sea bass. Handling of live fish, however, was only required for the tilapia study, since blood and fin-tissue had been previously collected from earlier experimental studies on halibut and sea bass respectively. Fish husbandry of the other species is elaborated in the relevant chapters. Nile tilapia (*Oreochromis niloticus*) was introduced to the University of Stirling in 1979 from Lake Manzala, Egypt. The basic maintenance of the experimental stock followed working procedures under ASPA (Animals Scientific Procedures Act, 1986) and monitored by the Home Office in the United Kingdom. An accredited training for personnel working under ASPA had to be followed and a Personal Licence obtained before carrying out experimental work with fish approved by ASPA (PPL 60/4397).

Individual brood fish were held in square fibreglass tanks (generally the males) or in glass aquaria (generally the females). Continuous water flow and aeration was in operation in each tank. All tanks were in recirculating systems within a controllable rearing environment and with proper facilities for filtering and cleaning the water before recycling back through the fish rearing tanks. New water was added to the systems to replace the water used to flush out fish waste materials as well as to make up water loss due to evaporation. The water quality parameters, particularly dissolved oxygen, ammonia and nitrate and nitrite contents were checked on a

weekly basis. The standard temperature of the water in the Tropical Aquarium Facility (TAF) was maintained at around 28^{o}C.

Handling of fish was performed with care. Nets of proper mesh sizes were used to capture the fish, which were then held in plastic buckets filled with water from storage tanks (temperature 28 ^{o}C) before being transferred to the wet laboratory for experimental procedures. Nets were soaked in disinfectants (Total Farm Iodophor; major components phosphoric acid, iodine and non-ionic surfactant**)** before and after use. Fish were anesthetised to avoid excessive handling stress prior to tagging, breeding or fin biopsy sampling. For this purpose, benzocaine (ethyl-4-aminobenzoate, Sigma-Aldrich, UK) solution at a final concentration of 1:10,000 was used. A stock solution was first prepared by dissolving benzocaine powder at 10% (w/v) in ethanol.

Brood fish were tagged by a TROVAN Passive Integrated Transponder (PIT) tag that had a unique 10-digit code. The fish were anesthetised and tagged with the aid of a special wide tip syringe (previously disinfected in 70% ethanol) on the lateral-abdominal side of the fish, lifting a scale and making an incision under it. The incision was sealed with the same scale and fish was immediately placed into clean aerated water until full recovery before return to the original tank or an aquarium to rear as broodstock.

Females were kept in glass aquaria so that they could be easily observed for signs of readiness to spawn. A swollen reddish genital papilla was the main sign of this.

Stripping was carried out in the wet lab of the TAF. The anesthetised fish was taken out of the bucket and placed on the bench on wet tissue paper. The eggs were released by slight ventral pressure (stripping) and kept in a Petri dish partially filled with clean water (temperature 28 °C) directly from the incubator system where the eggs were to be hatched. The spent fish was immediately transferred to another bucket filled with oxygenated water for recovery. The male fish was then placed in anesthetic solution. Meanwhile, the collected eggs were washed several times with aquarium water to remove any faeces, mucus and scales and were held in enough water just to cover them. The milt was collected from anaesthetised selected males in glass capillaries (1 mm diameter, Drummond Scientific Co. USA) directly from the urogenital pore of the male once the urine was drained, by applying ventral pressure. This avoided activation of the sperm. Milt was added gently to the eggs, stirred and left standing for 2-3 minutes to ensure maximum fertilization rate before the eggs were transferred into a plastic down-welling incubator in a recirculation system. The spent fish were transferred to an aerated bucket filled with water for about 5 minutes for recovery and then returned to their original tanks.

Fertilized eggs were washed with aquarium water and transferred to a series of 750 ml round bottomed plastic jars (soft drink bottles) for incubation. These jars were connected to a recirculating system where warm water was fed from a 125 litre (L) overhead tank to the jars by gravity. The water from the overhead tank first passed through a 30 W UV sterilisation unit (flow rate 20 L/min, UV dosage 62,000 μW cm^2), then through 20 mm PVC pipe to the jars. They received water from the PVC pipe flow via 4-m diameter Perspex tubing connection and the flow in the jars was

controlled by small airline taps in such a way that the eggs in the jars were kept in gentle motion at all times. The wastewater was discharged into the biofiltration tank (180 L capacity) via two filters filled with fibre wool positioned just above the settling tank. The shell filters helped to maintain the pH of the system and act as a surface for bacteria. Fertilized eggs were kept in hatching incubators for a total period of 10 days before they were ready to transfer to circular plastic tanks. The initial number of eggs was recorded and dead eggs and embryos were removed by siphoning during this period.

Fry were then transferred into circular or rectangular plastic tanks in a recirculation system and reared for 3-4 months. The fry density was 50-80 per tank. After this period the fish were immersed in an overdose of anaesthetic (0.05% benzocaine) for about 10 minutes before destruction of the brain and being sexed by gonad dissection and examination of a squash preparation. Fin samples were then stored in 100% ethanol tubes for subsequent DNA extraction.

2.1.2 Temperature treatment

Fry from a single family at 10 dpf were split into two groups. A control group underwent the procedures described in previous section and a treatment group were placed into plastic aquarium containing 5 l of water at 28^{o}C. Aquariums were then placed into heated trout egg trough at approx. 37^{o}C (fine tuned to reach 36.0^{o}C inside 5 l aquaria). The temperature inside aquarium equilibrated to 36.0^{o}C in about 2.5 hr. Removal of waste by siphoning and water replacement was performed twice a day

(daily/as required; normally 0.5-1.0 l; water replaced from aquarium P10 in heated trout egg trough). After 10 days of heat treatment, fish plus 5 l of water from aquarium were transferred into an empty 20 l tanks in a recirculation system. Water flow was adjusted so that temperature would equilibrate at 28^{o}C in about 2.5-3 hr. Subsequent rearing and sexing was performed as previously described.

2.1.3 Feeding regimes

Fish of all sizes (from fry to broodstock) were fed with an appropriate size (designated as no. 3-5) of commercial trout feed (Trouw Aquaculture Nutrition, UK; manufacturer Skretting, Preston, UK; Product code 470405) twice a day. Hatchlings were fed with powdered food (0.25-0.50 mm in diameter), prepared by grinding feed size 5 (4 mm diameter) and fed *ad libitum* for the first four to six weeks followed by feeding a mixture of ground food and no. 3 feed size as the fish grew. Advanced fry and fingerlings weighing between 5 g to 40 g received no. 3 size feed twice a day at a rate of approximately 5% of their body weight. Fish weighing 40 g to 80 g and >80 g were fed with no. 4 and no. 5 sized feed respectively, at a rate of approximately 2% of their body weight per day.

2.1.4 Composition of feed ingredients

The feed contained the following ingredients: Wheat, Soybean meal, maize gluten (60%), fishmeal, sunflower meal, fish oil, minerals, vitamins

2.1.5 Nutrient compositions

Nutrient composition was as followed: Trout food no.3: 8% oil, 5.7% ash, 0.9% phosphorus, 38% protein, 4% fibre, +12000 iu/kg Vit.A, +2000 iu/kg Vit.D3, 100 iu/kg Vit.E as alpha tocopherol acetate, antioxidants (BHT, butylated hydroxytoluene; BHA, Butylated hydroxyanisole).

2.2 DNA extraction protocol

DNA was extracted using the REALpure kit (REAL Laboratories Spain). Fin clip (approximately 20 mg) or blood samples (approximately 20 μl) were placed in individual nucleic acid free 1.5 ml tubes and 200 μl of lysis solution was added. Five μl of proteinase K (10 ul/ul) were added, mixed with gentle vortexing for 1 min and incubated overnight at 55 ^{o}C. Five μl of RNAse (10 ul/ul) were added to each tube and gently vortexed for 1-2 min followed by incubation at 37 ^{o}C for 60 min. Samples were brought to room temperature and 100 μl of protein precipitation solution was added, mixed and then centrifuged at 21,000 rcf for 10 minutes. Precipitated protein formed a pellet. 180 μl of the supernatant was poured into a new tube containing the same volume of isopropanol and mixed a few times. The tubes were centrifuged at 14,000 rcf for 1 min and the supernatant was poured off. Two ethanol washes were performed. 400 μl of 75% ethanol was added, and then the tube was centrifuged at 14,000 rcf for 1 min. Ethanol was poured off and the same procedure was repeated once more. The samples were air-dried by keeping upside down on absorbent paper for approximately 1 hour. 30 μl of 5mM Tris (pH 8.5) was added to each tube. Each sample was quantified by spectrophotometry (Nanodrop) and quality of DNA was

assessed by agarose gel electrophoresis, and each sample was finally diluted to a concentration of 50 ng/μL in 5 mM Tris, pH 8.5.

2.3 RAD library preparation and sequencing

The RAD library preparation protocol followed essentially the methodology originally described in Baird et al. (2008) and comprehensively detailed in Etter et al. (2010) with the minor modifications described in Houston et al. (2012). The RAD specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. (2011). Parental DNA was used in excess to ensure accurate coverage of their genotypes.

Each sample (0.72 μg parental DNA / 0.24 μg offspring DNA) was digested at 37°C for 40 minutes with *Sbf*I (recognising the CCTGCA|GG motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U *Sbf*I per μg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 μg DNA per 50 μL reaction volume. The reactions (12 μL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual specific P1 adapters, each with a unique 5 bp barcode, were ligated to the *Sbf*I digested DNA at 22ºC for 60 minutes by adding 1.8 / 0.6 μL 100 nmol/L P1 adapter, 0.45 / 0.15 μL 100 mmol/L rATP (Promega), 0.75 / 0.25 μL 10× Reaction Buffer 2 (NEB), 0.36 / 0.12 μL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 45 / 15 μL with nuclease-free water for each parental / offspring sample. Following heat inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in appropriate multiplex pools.

Shearing (Covaris S2 sonication) and initial size selection (250-700 bp) by agarose gel separation was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation, library amplification, exactly as in the original RAD protocol. A total of 150 µL of each amplified library (14-18 PCR cycles, library dependent) was size selected (c. 350-650 bp) by gel electrophoresis. Following a final gel elution step into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing using the HiSeq 2000 Illumina platform.

2.4 ddRAD library preparation and sequencing

The ddRAD library preparation protocol followed a modified version of the methodology originally described in Peterson et al (2012). Each sample (0.72 µg parental DNA / 0.24 µg offspring DNA) was digested at 37°C for 40 minutes with *Sbf*I (recognising the CCTGCA|GG motif) and *SphI* (recognising GCATG|C motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U each enzyme per µg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 µg DNA per 50 µL reaction volume. The reactions (12 µL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual-specific combinations of P1 and P2 adapters, each with a unique 5 bp barcode, were ligated to the digested DNA at 22ºC for 60 minutes by adding 1.8 / 0.6 µL 100 nmol/L adapters, 0.45 / 0.15 µL 100 mmol/L rATP (Promega), 0.75 / 0.25 µL 10× Reaction Buffer 2 (NEB), 0.36 / 0.12 µL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 45 / 15 µL

with nuclease-free water for each parental / offspring sample. Following heat inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in a single pool (for one sequencing lane) and purified. Size selection (300-600 bp) was performed by agarose gel separation and followed by gel purification and PCR amplification. A total of 100 µl of the amplified library (13 cycles) was purified using an equal volume of AMPure beads. After eluting into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing.

2.5 SNP assays

We designed SNP assays using the KASP genotyping system (KBioscience UK Ltd). Allele-specific primers and other assay components were supplied by KBioscience UK Ltd, based on the supplied marker sequences (Table S3, Data S1). PCR reactions were carried for 10 µL final volume reactions. The cycling conditions were the following: 94ºC for 15 min, 94ºC for 20 sec, touchdown over 61ºC to 55ºC for 60 sec (10 cycles dropping 0.8ºC each cycle) and an extra 34 cycles at 55ºC.

2.6 Data analysis

2.6.1 General statistics -Parentage assignment

R v.3.0.1 was used for calculating descriptive statistics, conducting chi-square tests for significant deviations from equal sex ratios and running both general and generalized linear models. Parentage assignment was performed using Vitassign V8-5.1 (Vandeputte et al., 2006) using 200 SNPs.

2.6.2 Genotyping RAD alleles

Reads of low quality (score under 30), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotypes using Stacks software (Catchen et al., 2011). The likelihood-based SNP calling algorithm (Hohenlohe et al., 2011) implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. The parameters were a minimum stack depth of at least 30, a maximum of 2 mismatches allowed in a locus in an individual and up to 1 mismatch between alleles.

2.6.3 Genetic map construction

The genetic maps were constructed using R/Onemap (Margarido et al., 2007) and TMAP (Cartwright et al., 2007). SNPs were initially tested for segregation distortion

using the *chisq* module of TMAP. Recombination rates and allocation of markers into linkage groups was conducted using R/Onemap (functions: *rf.2pts, group*). This package uses Hidden Markov Models (HMM) algorithms for outbred species while in parallel implements the methodology described in Wu et al. (2002), for calculating the most probable linkage phase. Linkage groups were formed using a minimum LOD value of 6-10. Both R/Onemap (functions: *ug, rcd* and *record*) and TMAP were used to order the markers in every linkage group. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function. The genetic maps were drawn and aligned using Genetic-Mapper v0.3 (Bekaert, 2012).

2.6.4 QTL mapping

QTL analysis was performed using different software, like R/qtl, GridQTL and QTLMap, in order to strengthen the validity of the results. Models based on both single and multidimensional approach were used (R/qtl functions: *addqtl*, *scantwo, fitqtl, stepwiseqtl,* GridQTL and QTLmap: scanning for two QTL simultaneously). Permutation tests (10,000 permutations) were conducted in order to correct for the multiple testing.

2.6.5 Association analysis

Association analysis was performed using R/GenABEL (Aulchenko et al., 2007) and SNPassoc (Gonzalez et al., 2007) in order to test for existing SNPs associated with sex. The first software was also used in estimating heritabilities. Both the Bonferroni and permutation tests (10,000 permutations) were used in order to correct for multiple testing.

Chapter 3

Atlantic halibut

3 Mapping the sex determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing

Christos Palaiokostas [1], Michaël Bekaert [1], Andrew Davie [1], Mairi E. Cowan [1], Munevver Oral [1], John B. Taggart [1], Karim Gharbi [2], Brendan J. McAndrew [1], David J. Penman [1], Hervé Migaud [1]

[1] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2] The GenePool, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK

Keywords: *Hippoglossus hippoglossus*, Sex determination, Mono-sex, QTL Mapping, RAD-seq, Aquaculture.

Abbreviations: RAD: restriction-site associated DNA; SNP: single nucleotide polymorphism; QTL: Quantitative trait locus; MAS Marker assisted selection; LG: Linkage group.

Status: *BMC Genomics 14:556*

Contribution: The first draft of the present manuscript was compiled and written in full by the author of this thesis, who was also fully involved in all subsequent revisions. DNA extraction, preparation of RAD libraries (under guidance of John Taggart), construction of the linkage map, QTL mapping, association analysis and general statistics were carried by the candidate. The other co-authors contributed towards the experimental design, sexing the fish, the analysis of sequenced reads, genotyping for the SNP assays, synteny analysis and combined marker sex prediction.

Abstract

Atlantic halibut (*Hippoglossus hippoglossus*) is a high-value, niche market species for cold-water marine aquaculture. Production of mono-sex female stocks is desirable in commercial production since females grow faster and mature later than males. Understanding the sex determination mechanism and developing sex-associated markers will shorten the time for the development of mono-sex female production, thus decreasing the costs of farming. Halibut juveniles were masculinised with 17α-methyldihydrotestosterone (MDHT) and grown to maturity. Progeny groups from four treated males were reared and sexed. Two of these groups (n = 26 and 70) consisted of only females, while the other two (n = 30 and 71) contained balanced sex ratios (50% and 48% females respectively). DNA from parents and offspring from the two mixed-sex families were used as a template for Restriction-site Associated DNA (RAD) sequencing. The 648 million raw reads produced 90,105 unique RAD-tags. A linkage map was constructed based on 5,703 Single Nucleotide Polymorphism (SNP) markers and 7 microsatellites consisting of 24 linkage groups, which corresponds to the number of chromosome pairs in this species. A major sex-determining locus was mapped to linkage group 13 in both families. Assays for 10 SNPs with significant association with phenotypic were tested in both population data and in 3 additional families. Using a variety of machine-learning algorithms 97% correct classification could be obtained with the 3% of errors being phenotypic males predicted to be females. Altogether our findings support the hypothesis that the Atlantic halibut has an XX/XY sex determination system. Assays are described for sex-associated DNA markers developed from the RAD sequencing analysis to fast track progeny testing and implement monosex

female halibut production for an immediate improvement in productivity but also speed-up the inclusion of neomales derived from many families to maintain a larger effective population size and ensure long-term improvement through selective breeding.

3.1 Introduction

The mechanisms of sex determination in animals are remarkably diverse. Gonochoristic animals show genetic and/or environmental sex-determining mechanisms. Genetic sex-determining systems can be either chromosomal, and involve a master sex-determining gene/region on a sex chromosome, or can be polygenic and involve several genes/regions on multiple chromosomes. In most fish species with XX/XY or ZZ/ZW mechanism, the sex chromosomes do not show clear differences in length or gene content (Devlin & Nagahama, 2002). Several fish sex determining genes have been isolated from species with XX/XY mechanisms: *DMY/Dmrt1bY* in medaka (*Oryzias latipes*) (Matsuda et al., 2002); *Gsdf(Y)* in Luzon ricefish (*Oryzias luzonensis*) (Myosho et al., 2012); *amhy* in Patagonian pejerrey (*Odontesthes hatchery*) (Hattori et al., 2012); *Amhr2* in tiger pufferfish (*Takifugu rubripes*) (Kamiya et al., 2012); and *sdY* in rainbow trout (*Oncorhynchus mykiss*) (Yano et al., 2012). In environmental sex-determining systems, the environment plays a decisive role, such as temperature in turtles, alligators and fish (Devlin & Nagahama, 2002; Bull & Vogt, 1979; Barske & Capel, 2008). Both systems can interact in some species such as in medaka, which has an XX/XY genetic system, where high temperatures can cause female-to-male sex reversal (Nanda et al., 2002, Sato et al., 2005; Barske & Capel, 2008). Additionally, autosomal loci can also contribute to sex determination in many species (Lee et al., 2004). Overall, the

understanding of sex determination systems in fish has direct commercial applications, given the strong sexual dimorphism exhibited in a wide variety of aquaculture fish species for a range of commercially important traits like growth or age at maturity.

Atlantic halibut (*Hippoglossus hippoglossus*) has been a high-value species for cold-water marine aquaculture for several decades in Northern Europe and America, although production has been limited by a series of bottlenecks. Among these, sexual dimorphism in growth, with males maturing earlier and growing significantly slower than females, reduces productivity and profitability of the sector. Females can reach market size (3-5 Kg) at around 36 months while males require at least an extra year, making the production of all-female stocks particularly appealing for the aquaculture industry (Bjornsson, 1995; Babiak et al., 2012).

Flatfish (order Pleuronectiformes) show a range of sex-determining mechanisms, including XX/XY and ZZ/ZW, with significant effects of environmental factors, principally temperature, in some species (Luckenbach et al., 2009). Meiotic gynogenetic *H. hippoglossus* were all female, suggesting an XX/XY sex-determining system (Hendry et al., 2002). Temperature has not been shown to have an effect on *H. hippoglossus* sex ratio (Hughes et al., 2008). Gonadal sex differentiation can be manipulated through in-feed synthetic steroid treatments (*e.g.*, 17α-methylhydrotestosterone, MDHT or 17β-estradiol) (Hendry et al 2003) or aromatase inhibitor treatments (*e.g.*, Fadrozole) (Babiak et al., 2012). However, direct sex reversal is not a commercially acceptable means to alter sex ratios in food fish within

the EU (Directive 2003/74/EC). Thus indirect sex reversal is required, whereby masculinised genotypic females (XX neomales) are crossed to normal females (XX) to produce genetically all-female progeny, a process which has yet to be proven in *H. hippoglossus*. The crux of successful indirect sex reversal is the non-lethal identification of the neomales. Currently the main technique for such verification is progeny testing of treated animals which is time consuming and costly, taking at least four or five years due to the timing of puberty in halibut (reached after three years). Direct genetic sexing, instead of progeny testing, would be preferable using non-lethal and cheap genotyping techniques. This is only likely to be possible in simple cases of male or female heterogamety. Sex-specific genomic sequences are only available in a limited number of aquaculture species (Piferrer & Guiguen, 2008). Although a genetic linkage map based on microsatellites and amplified fragment length polymorphism (AFLP) is available for *H. hippoglossus* (Reid et al., 2007), this does not contain any information about sex-determination. Restriction-site associated DNA (RAD) sequencing is a powerful technique for generation of high-density linkage maps and conducting quantitative trait locus (QTL) analysis (Miller et al., 2007; Baird et al., 2008) including the mapping of sex-determining loci in fish (Anderson et al., 2012).

The aim of the current research was to demonstrate that indirect sex reversal was possible and thereafter to develop sex-associated markers through RAD-sequencing. An in-feed MDHT treatment was given to weaned halibut juveniles during the labile period, which resulted in 97% phenotypic males. A sub-population of these treated fish was then reared to maturity and from this stock, two neomales and two normal

males were verified by progeny testing, a process that took four years to complete. The sex-determining locus was mapped to the end of the linkage group 13, in the two mixed sex families from the sex reversal study, using polymorphic Single Nucleotide Polymorphisms (SNP). A combination of four markers predicted sex with 97% accuracy in any individual fish, from a panel of progeny and broodstock. Synteny analysis showed that DNA sequences containing Atlantic halibut sex-associated SNPs were consistently clustered in several other fish genomes. These results suggested that sex determination in Atlantic halibut is likely to be monogenic (XX/XY) and localised within a 3.2 cM window on linkage group 13.

3.2 Materials and Methods

3.2.1 Hormonal sex reversal

Weaned mixed-sex halibut larvae (mean total length of 40.1 ± 0.2 mm, mean wet weight of 0.5 ± 0.01 g) produced in the 2007 spawning season, were obtained from a commercial halibut hatchery and transferred to the Machrihanish Marine Environmental Research Laboratory (55.424°N, 5.749°W) for hormonal treatment. Three in-feed treatments were tested in duplicate: a) 6 weeks steroid free diet (control), b) 6 weeks MDHT in-feed (5 ppm) and c) 3 weeks MDHT in-feed (10 ppm) followed by 3 weeks steroid-free diet. Food was provided in excess by automated feeders into the tanks every 12 minutes throughout a 24-hour period. Feed, based on a commercial diet (Low Energy Marine Larval diet, EWOS, West Lothian, UK), was mixed with an ethanol solution containing the appropriate dose of

MDHT (Sigma-Aldrich Co Ltd, Poole, UK) and then dried in an extraction fume hood. Following treatment and once the fish had reached a mean weight of 28.4 ± 0.4 g, replicate treatment groups were identified by a coded subcutaneous dye mark and then reared communally. At approximately 1-year post fertilisation a total of 80 individuals per treatment (40 per replicate) were sacrificed and fixed in 4% neutrally buffered formalin for histological determination of phenotypic sex. Sex ratios were compared to the expected 1:1 and were evaluated statistically using a chi-square test.

3.2.2 Neomale verification by progeny testing

At a mean size of 180.8 ± 3.1 g, 60 control fish (30 per replicate) and 150 fish from the 5 ppm treatment (75 per replicate) were tagged with a passive integrated transponder tag (Fish Eagle Co., Lechlade, UK). Fish were then reared communally until first maturity in spring 2010. In March 2010, crosses were performed between 7 males from the hormone-treated population and normal female broodstock. Fertilisation was confirmed in each cross by microscopic examination of blastomere development. Eggs from each cross were maintained in isolation using standard commercial rearing methodologies. Sufficient progeny from only four of these males survived through yolk sac absorption, live feeding and weaning. These four families were reared in isolation at a commercial halibut hatchery until phenotypic sex ratio could be assessed in February-March 2011, once fish reached a suitable size (> 50 g) for histological sexing of the gonads. A total of 30 (family A & B) or 70 (family C & D) individuals/family were sacrificed for histological examination and blood was

sampled for genotyping (total of 200 offspring). Sex ratios were compared to the expected 1:1 using chi-square test.

3.2.3 RAD library preparation and sequencing

DNA was extracted from blood samples of the fish using the REALPure genomic DNA extraction kit (Durviz S.L.) and treated with RNase to remove residual RNA from the sample. Each sample was quantified by spectrophotometry (Nanodrop) and quality assessed by agarose gel electrophoresis, and was finally diluted to a concentration of 50 ng/μL in 5 mmol/L Tris, pH 8.5. The RAD library preparation protocol followed essentially the methodology originally described in Baird et al. (2008) and comprehensively detailed in Etter et al. (2011), with the minor modifications described in Houston et al (2012) . The RAD-specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. (2011).

Each sample (1.5 μg parental DNA / 0.5 μg offspring DNA) was digested at 37°C for 30 minutes with *Sbf*I (recognising the CCTGCA|GG motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U *Sbf*I per μg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 μg DNA per 50 μL reaction volume. The reactions (75 / 25 μL final volumes for parental / offspring samples respectively) were then heat inactivated at 65°C for 20 minutes. Individual specific P1 adapters, each with a unique 5 bp barcode (Table 1), were ligated to the *Sbf*I digested DNA at 22ºC for 45 minutes by adding 3.75 / 1.25 μL 100 nmol/L P1 adapter, 0.9 / 0.3 μL 100 mmol/L rATP (Promega), 1.5 / 0.5 μL 10×

Reaction Buffer 2 (NEB), 0.75 / 0.25 μL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 90 / 30 μL with nuclease-free water for each parental / offspring sample. Following heat inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in appropriate multiplex pools. Shearing (Covaris S2 sonication) and initial size selection (250-500 bp) by agarose gel separation was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation, library amplification, exactly as in the original RAD protocol (Baird et al 2008, Etter et al 2011). A total of 150 μL of each amplified library (14 PCR cycles) was size selected (c. 300-550 bp) by gel electrophoresis. Following a final gel elution step into 20 μL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent at the GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in one lane of an Illumina HiSeq 2000 using 100 base paired-end reads (v3 chemistry). Raw reads were process using RTA 1.12.4.2 and Casava 1.6 (Illumina). The reads were deposited at the NCBI BioProject under the accession SRP016043.

3.2.4 Genotyping RAD alleles

Reads of low quality (score under 30, while the average quality score was 37), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotyped using Stacks software 0.9995 (Catchen et al., 2011). The likelihood-based SNP calling algorithm (Hohenlohe et al., 2010) implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. The

parameters were a minimum stack depth of at least 30, a maximum of 2 mismatches allowed in a locus in an individual and up to 1 mismatch between alleles. The pair-ends were assembled using Stacks and Velvet version 1.2.08 (Zerbino & Birney, 2008) and used to separate RAD-tag sequence with or without potential SNP but belonging to separate loci (duplication products). Polymorphic RAD-tags may contain more than one SNP, but the vast majority (over 99%) showed only two allelic versions.

3.2.5 Genetic map construction

The genetic map was constructed using R/Onemap (Margarido et al., 2007) and TMAP (Cartwright et al., 2007). The allocation of markers in linkage groups was conducted using R/Onemap. This package uses Hidden Markov Models (HMM) algorithms for outbred species while in parallel implements the methodology described in Wu et al. (2002) for calculating the most probable linkage phase. Linkage groups were formed using minimum LOD values of 10. TMAP was used initially to phase the markers in every linkage group and then to order them. By using an HMM maximum likelihood model and taking into account potential genotypic errors it reduces the tendency to erroneously derive oversized linkage groups, a phenomenon which is often observed in dense maps (Cartwright et al., 2007). Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function. The genetic map was drawn and aligned using Genetic-Mapper v0.3 (Bekaert, 2012).

3.2.6 QTL- Association mapping

The QTL analysis was performed using three different suites of programmes: R/qtl (Broman & Sen, 2009), GridQTL (Seaton et al., 2006) and QtlMap (Gilbert et al., 2008). In the case of R/qtl the genotypes the two families were analysed separately. The analysis was performed considering the cross as a 'pseudo' backcross, effectively analysing male and female informative markers separately. The model used for the analysis was based on Interval Mapping. The phenotype was considered a binary trait (0 for females and 1 for males). The algorithm used considers the phenotype to follow a mixture of Bernoulli distributions and uses a form of the EM algorithm for obtaining maximum likelihood estimates (Broman & Sen, 2009). Two-way and multiple QTL models were also run with this package.

Approximate Bayesian and 1.5-LOD 95% density and confidence intervals were calculated respectively. An approximate estimate of the phenotypic variance explained by the QTL was obtained from the following equation: $1\text{-}10^{-2LOD/n}$. While the estimated variance may be reasonable for additive QTL, problems can be caused in the case of linked QTL (Broman & Sen, 2009). The GridQTL software was used to estimate the polymorphism information content across the genetic map. QTLMap was used for performing a joint QTL Analysis of the two families using SNPs from LG 13. The phenotype was considered as discrete and the model used was a Mixture Linkage Analysis model, accounting for heteroskedasticity.

An Association Analysis was performed for the two families using R/GenABEL (Aulchenko et al., 2007a) in order to identify SNPs associated with sex. The SNP data were tested for association using the fast score test for association (Aulchenko et al., 2007b). In all the above analysis genome-wide significance thresholds were calculated by permutation tests (10,000 permutations) in order to correct for multiple testing.

3.2.7 Verification of SNP-sex association

Marker sex association was tested using 10 competitive fluorescent, allele specific endpoint-genotyping assays (KASP v4.0, LGC genomics) based on SNPs that were commonly found in the two mapping families to span the region of highest association with sex. (*Hhi6696, Hhi7153, Hhi9493, Hhi10170, Hhi11772, Hhi18571, Hhi41238, Hhi47769, Hhi51454, Hhi58665*, NCBI dbSNP accession 749737483, 749737484, 749737485, 749737486, 749737487, 749737488, 749737489, 749737490, 749737491 and 749737492 respectively; Appendix: Additional file 3). SNP-specific primer sets were designed by LGC genomics (Appendix: Additional file 4). Each genotyping assay was run in an 8 μl volume containing approximately 40 ng of target gDNA incorporated with a proprietary reaction mix in accordance with the manufacturers guidelines.

All assays were run using the same touchdown thermal cycling programme as follows: 94°C for 15 minutes followed by 10 cycles of 94°C for 20 seconds melt, 65-57°C for 1 minute anneal and extension (decreasing of 0.8°C per cycle) followed by 26 cycles of 94°C for 20 seconds melt, 57°C for 1 minute anneal and extension.

There was one exception, SNP *Hhi*58665, for which the extension time was extended to 2 minutes. All assays were run in a Biometra TGradient thermal cycler (Biometra GmbH, Goettingen, Germany). Thereafter assays results were read at 25°C using an endpoint genotyping programme in a Techne Quantica qPCR thermal cycler (Bibby Scientific Ltd, Stone, UK) in which unknown genotypes were assigned based on fluorescent output in comparison to non-template control wells containing DNA/RNA free H_2O.

All 10 SNP assays were tested in 58 offspring from three halibut families produced in the commercial halibut hatchery, which were independent from the initial mapping families, and in 36 independent broodstock halibut (18 ♀:18 ♂) originating from the Shetland Isles, Iceland and the Faroe Islands were genotyped. An association analysis was performed using R/SNPassoc (González et al., 2007). In the case of family data, association was tested both in separate families and across all families together. A Bernoulli generalised linear model was applied in order to test the magnitude of association between the SNP genotypes and phenotypic sex using this package (function *association*). Both the Bonferroni and permutation tests (10,000 permutations) were used in order to correct for multiple testing.

3.2.8 Sex prediction

The KASP allele type of all markers for each individual tested along with their sex were entered into the WEKA package (Hall et al., 2009), which contains a variety of machine-learning algorithms, including JRip, an optimised rule learning algorithm. This classifier implements a propositional rule learner, Repeated Incremental

Pruning to Produce Error Reduction (RIPPER), which was proposed by Cohen (1995) as an optimised version of IREP. JRip builds additive rules based on the allele type of the markers. JRip then classifies each individual into a particular predicted sex based on the allele type of the markers for each individual. Permutatively, one individual was removed from the training set, and subsequently the algorithm then assigns its sex. The set of rules was stable between permutations (Figure 5).

3.2.9 Synteny searches

D. rerio, *G. morhua*, *G. aculeatus*, *L. chalumnae*, *O. niloticus*, *O. latipes*, *T. rubripes* and *T. nigroviridis* genomes were downloaded from Ensembl 68 (Flicek et al., 2012). We used BLASTN (Altschool et al., 1990) to perform a search for the RAD-tag (and their paired-ends) against the 8 fish genomes. The parameters used were minimum alignment size 80 nt, minimum percentage of sequence identity 0.25 and maximum e-value 0.001 and low complexity mask on. All other parameters were set as default to account for the divergence and shortness for the sequences used. Sequences that aligned to more than one place in each genome were excluded from further analysis.

3.3 Results

3.3.1 Hormonal sex reversal and neomale verification

The control group exhibited a sex ratio not significantly different from 1:1 (52% ♂; 48% ♀), whereas 97% of the group treated for six weeks (5 ppm) and 70% of the group treated for three weeks (10 ppm) were confirmed as phenotypic males (Table

1). Both in-feed treatments significantly altered the natural sex ratio in favour of males. Of the seven putative neomales that were progeny tested, only four crosses produced enough survivors at the age of sexing, at approximately one year of age. From these four crosses, two gave 100% female progeny (Families A and D; Table 1) while the other two gave balanced sex ratios (Families B and C; Table 3.1).

Table 3-1. Sex ratios in hormonal masculinisation trial (control, 5 ppm and 10 ppm MDHT) and progeny testing (families A-D from the four males from the 5 ppm MDHT group). Obs.; observed; Exp.: expected under H_0 hypothesis (1:1 ratio).

	Control	**5 ppm**	**10 ppm**	**A**	**B**	**C**	**D**
N	77	77	76	26	30	71	70
Male (Obs. / Exp.)	40 / 38.5	75 / 38.5	53 / 38	0 / 13	15 / 15	32 / 30.5	0 / 35
Female (Obs. / Exp.)	37 / 38.5	2 / 38.5	23 / 38	26 / 13	15 / 15	29 / 30.5	70 / 35
χ^2	0.12	69.21	11.8	26	0	0.15	70
***P*-value**	0.824	< 0.001	< 0.001	< 0.001	1	0.797	< 0.001
Sex Ratio	52% male	97% male	70% male	100% female	50% female	48% female	100% female

3.3.2 RAD sequencing

Two crosses with 62 (Family C) and 28 offspring (Family B) and their parents (including a common female) were analysed. The DNA samples were barcoded, pooled and sequenced in two lanes of an Illuminia HiSeq 2000 sequencer (Appendix: Chapter3: additional_files 1-4). In total, 648,402,546 raw reads (101 nt long) were produced (or 324,201,273 paired-ended reads: NCBI BioProject accession number SRP016043). After removing low quality sequences (quality score under 30), ambiguous barcodes and orphaned paired-end reads, 81.24% of the raw reads were retained (526,783,920 reads). The Stacks package (Catchen et al 2012) was then used

to make a *de-novo* assembly of the sampled loci from each individual: 82,745 and 83,668 RAD-tags (or RAD markers) were retrieved for Families B and C respectively, covering 90,105 RAD-tags in total including 76,308 of these shared between the two families (Figure 3.1). The number of reads and RAD-tags for each sample are reported in the Appendix: Additional file 1.

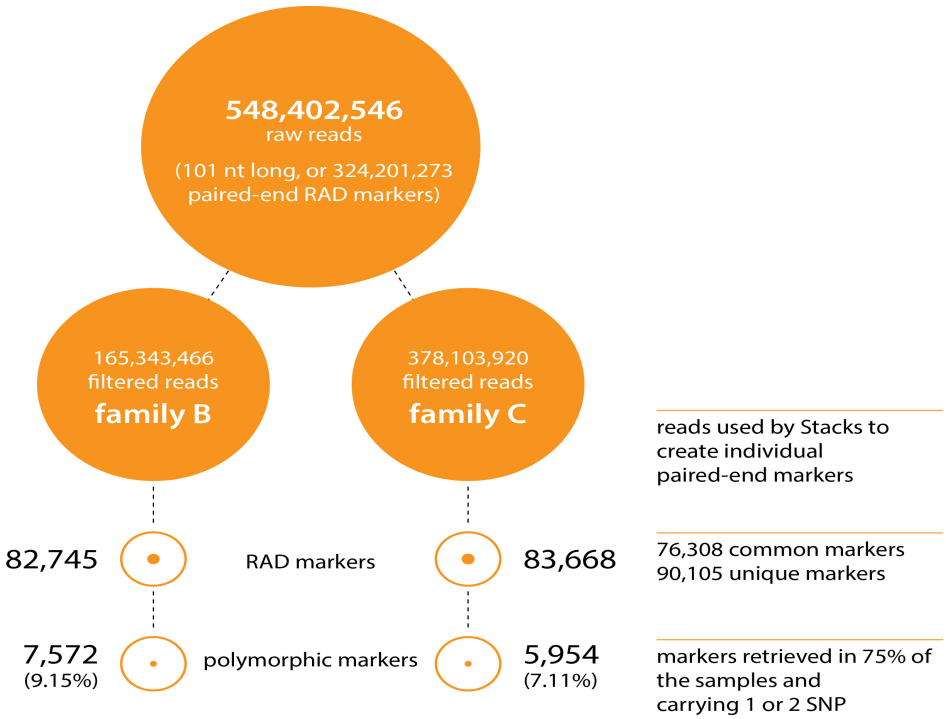

Figure 3-1. Sequencing and RAD-tag summary. Detail of the number of reads before and after filters (orange disk) followed by the reconstructed number of RAD markers and polymorphic RAD markers (orange circles).

3.3.3 Genetic Map

In order to maximise the number of informative markers and minimise the amount of missing or erroneous data, we used only paired-end RAD-tags retrieved in at least 75% of the samples in each family, and carrying one or two SNPs. 7,572 and 5,954

RAD-tags were retained for the families B, and C respectively (Figure 3.1). Since Family B had only 28 offspring, the genetic map was constructed with the Family C data only (62 offspring). The map consists of 5,703 SNPs and 7 microsatellites (used initially for parentage assignment) in 24 linkage groups (LG) and spanning 1,514 cM (Figure 3.2; Additional file 2). 4,049 of the above SNPs were common in the two families and were used to incorporate the data from Family B into a joint linkage map. Sex-specific genetic maps were constructed, with the female-specific map spanning 1,496 cM and the male-specific map spanning 1,378 cM (Table 3.2).

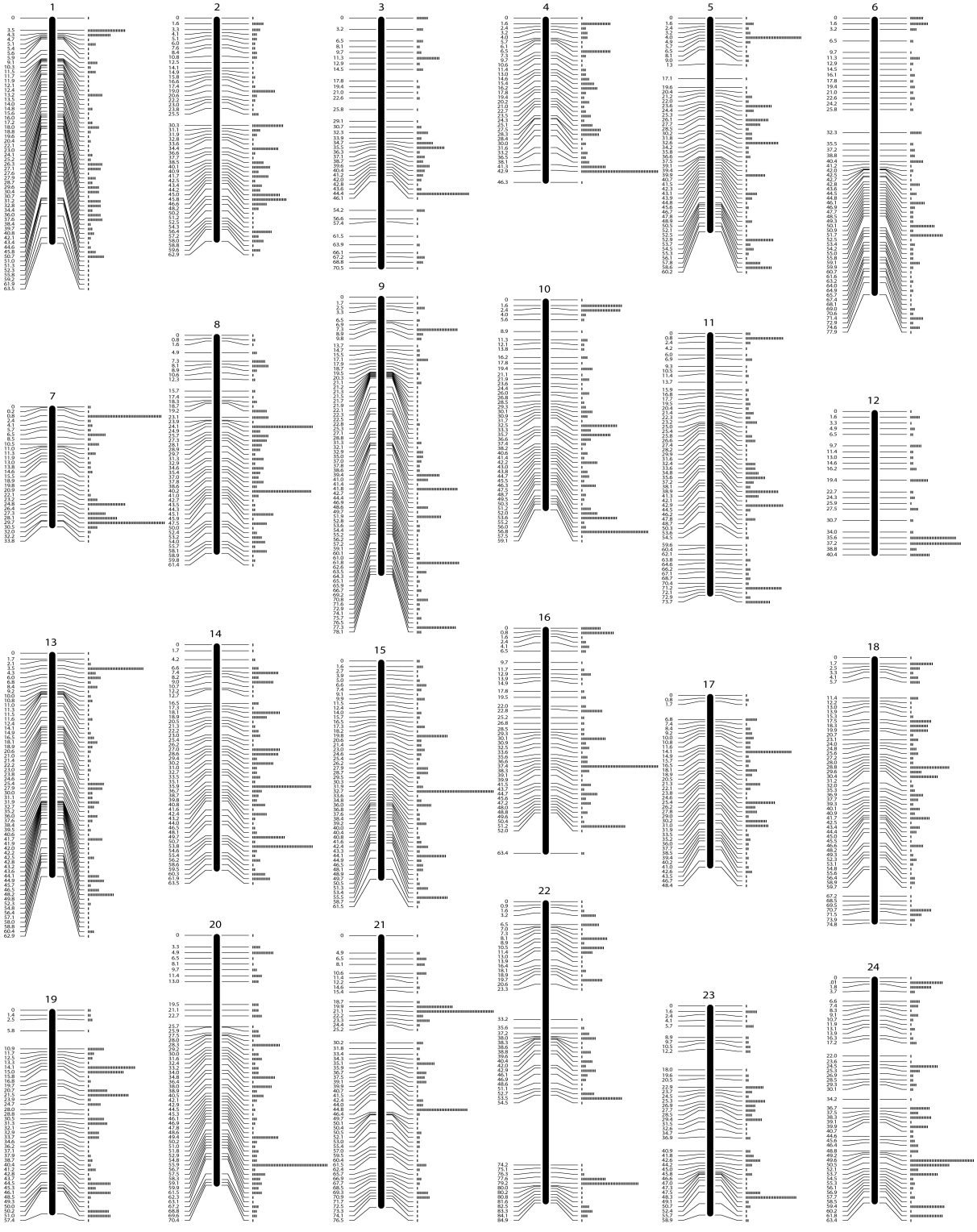

Figure 3-2. Genetic linkage map. Map with linkage group assignment determined using syntenic markers with previously published *H. hippoglossus* maps. The positions on the left side of the chromosomes are in cM. The rectangles on the right hand side represent the number of markers at this position. Detailed data is provided in the Appendix (Additional file 2).

Table 3-2. *H. hippoglossus* genetic map. Three linkage maps were constructed using all 5,710 segregating polymorphic markers from males, females or both.

Linkage Group	No. of markers	Size (cM)	Female (cM)	Male (cM)
1	215	63.5	60.0	63.5
2	240	62.9	62.9	58.0
3	177	70.5	67.2	38.2
4	229	46.3	44.7	28.4
5	241	60.2	58.6	58.6
6	196	77.9	74.6	37.5
7	230	33.8	32.0	33.8
8	268	61.4	61.4	61.4
9	265	78.1	78.1	78.1
10	293	59.1	57.5	57.5
11	246	73.7	73.7	73.7
12	186	69.2	40.4	64.3
13	222	62.9	62.9	60.8
14	318	63.5	61.9	63.5
15	269	61.5	61.5	59.9
16	237	63.4	63.4	63.4
17	174	48.4	48.4	48.4
18	257	74.8	73.1	74.8
19	250	57.4	57.4	51.0
20	269	70.4	70.4	44.7
21	260	76.5	76.5	66.0
22	250	84.9	84.9	81.7
23	169	58.9	57.3	49.1
24	303	63.4	63.4	61.8
Size (cM)		1,514	1,496	1,378.1
Markers	5,710		3,858	3,412

3.3.4 QTL-Association mapping

The results from the single-QTL model for binary traits provided evidence for the existence of a major QTL in LG 13 for both families in the male informative dataset (Figure 3.3A). The highest logarithm of odds (LOD) scores for Families B and C were 6.83 and 12.17 respectively (Figure 3B). The genome-wide thresholds ($\alpha = 0.01$) were calculated from permutation tests (10,000 permutations) to be 3.25 and 3.90 respectively. The highest LOD scores were observed in the region between 30 cM and 48 cM in LG 13 for both families. Even though models that take into account the existence of a major QTL (as in this study) or ones that test for existence of multiple QTLs simultaneously reduce the residual variation (providing this way higher power in the analysis for detecting additional QTLs at least of modest effect), no additional QTLs were detected.

The calculated 95% Bayesian Density Intervals for the QTL location spanned a region of 22 cM (28-50 cM in LG 13), while the 1.5-LOD support interval spanned a region of 28 cM (23-51 cM in LG 13). The variance explained by the QTL was around 82%. There were no significant QTL in the female informative dataset. The Polymorphism Information Content (PIC) of the markers in LG 13 ranged from 0.64 to 0.99. The joint QTL analysis of the two families showed a maximum at 46 cM with a Likelihood Ratio (LR) of 96 (Figure 3.3B). The chromosome-wide threshold calculated from permutation tests (10,000 permutations) was LR = 19. The association analysis identified 38 SNPs being strongly associated ($P < 0.001$) with

sex in the two families. Models scanning simultaneously for the existence of two QTL or multiple QTL did not reveal any extra QTL.

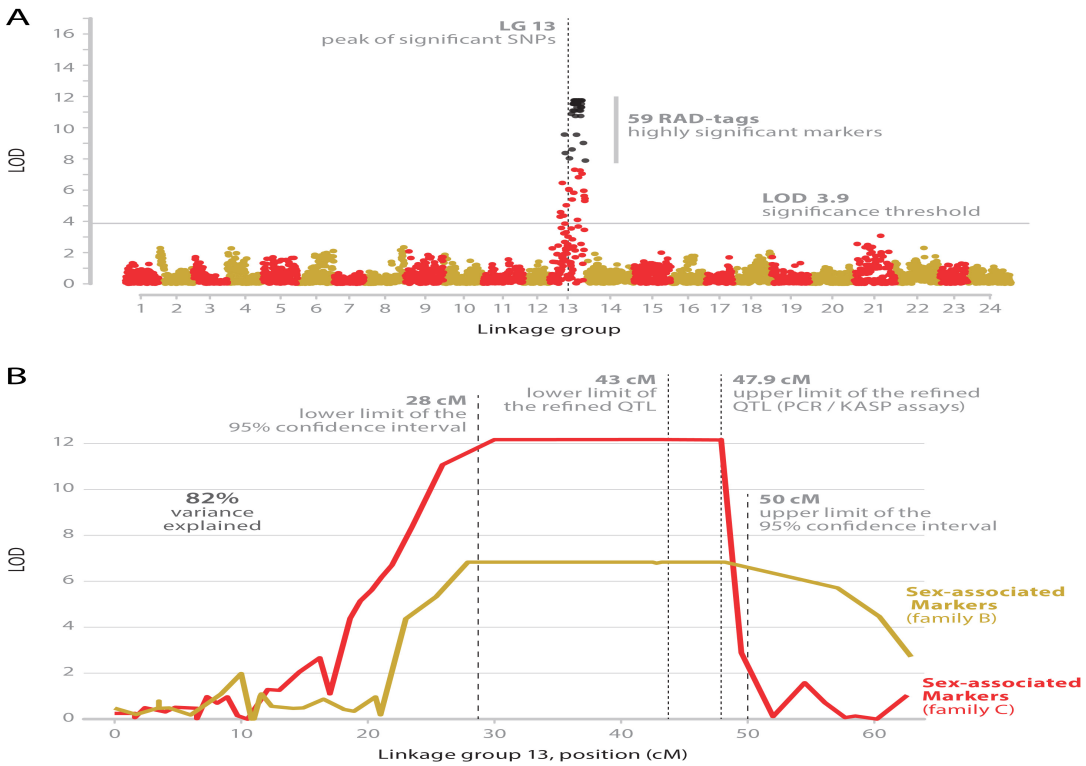

Figure 3-3. Results from QTL-Association analysis in Atlantic halibut families. (A) Association results for genotyped SNPs. SNPs with *p*-values achieving genome-wide significance ($P < 7.2 \times 10\text{-}8$) are shown in black (R/GenABEL). (B) Regional analysis of the QTL on LG 13. Plot of the LOD score (sex-association QTL search) along the linkage group 13 for family B and C (QTLMap).

3.3.5 Verification of SNP sex association

Sex association of 10 SNPs (Additional file 3) selected from LG 13 was investigated using allele specific endpoint-genotyping assays (Figure 3.4; Appendix: Additional file 3). The *p*-value thresholds ($\alpha = 0.05$) after taking into account multiple testing (for the 10 SNPs that were tested) according to the permutations and the Bonferroni

correction tests were found to be 0.0054 and 0.0050 respectively. *Hhi*58665 ($P = 3.05\times10^{-19}$), *Hhi*10170 ($P = 3.05\times10^{-19}$), *Hhi*41238 ($P = 2.45\times10^{-16}$) and *Hhi*47769 ($P = 8.96\times10^{-17}$) showed the highest association with sex, over 97% in the three test families (n = 10 ♀:10 ♂ in Families 1 & 2; 9 ♀:9 ♂ in Family 3). In each case the parents were a heterozygous ♂ and a homozygous ♀ for each marker. Furthermore when tested in wild sourced broodstock from locations across the species native range (Appendix: Additional file 4), the same four markers individually showed between 78% and 89% association with sex ($P < 0.001$), with SNP *Hhi*58665 showing the highest association ($P = 6.39\times10^{-7}$).

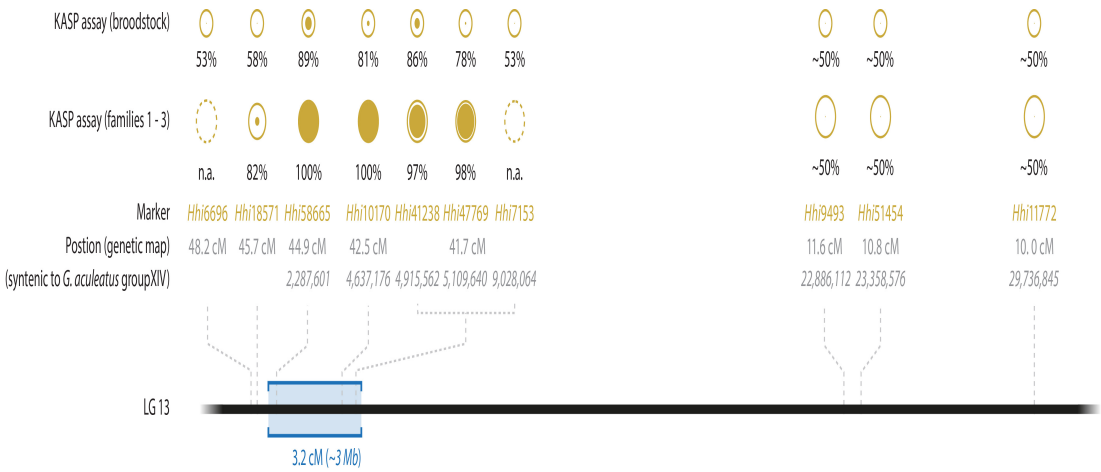

Figure 3-4. KASP assay and fine gene mapping on LG 13. Details of the 10 markers tested by KASP assay. From bottom to top: Location of the 10 markers (in the genetic map in cM and syntenic loci on the *G. aculeatus*, three-spined stickleback, group XIV in bp); KASP assay results. The outer circle diameters for the KASP assay results are proportional to the number of alleles tested. The inner (solid) disks represent the marker association with the phenotypic sex. Detailed data is provided in the Appendix (Additional file 4). When no informative polymorphism was found, "n.a." is specified

The combined prediction power of these four markers was tested on the 36 broodstock and 58 progeny using the JRip classifier, as the derived rules have straightforward interpretation. The combined prediction is based on two rules using three markers (Figure 3.5) and produced 97% corrected classification with the 3% of errors being phenotypic males predicted to be females. The male prediction precision is 1 (recall of 0.935) and the female prediction precision is 0.94 (recall of 1).

Within the 18 tested male broodstock, two had "female" genotypes for all four of these markers (one of the 58 progeny also had male phenotype but female genotype). One of these two broodstock had previously been crossed with four females and had produced only female offspring (between 3 and 14 individuals per family, total 27). The phenotypic sex of these 27 offspring was verified by post-mortem examination three years post-fertilisation.

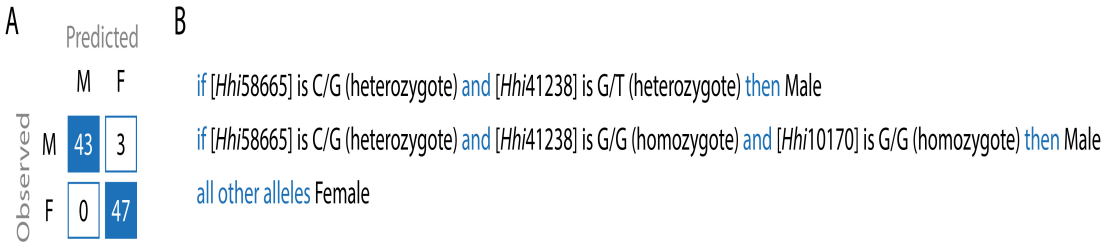

Figure 3-5. Combined marker sex prediction. (A) Confusion matrix of the JRip rules. Blue cells are correct predictions; white cells are the erroneous predictions. Overall the predictions are 97% accurate. (B) JRip rules based on the alleles detected using the KASP assays.

3.3.6 Synteny searches

We selected the 59 markers within the 95% confidence interval around the LOD score peak and mapped them onto the genomes of related species to identify syntenic regions. We performed this search against zebrafish (*Danio rerio*), Atlantic cod (*Gadus morhua*), three-spined stickleback (*Gasterosteus aculeatus*), West Indian ocean coelacanth (*Latimeria chalumnae*), Nile tilapia (*Oreochromis niloticus*), medaka (*Oryzias latipes*), tiger pufferfish (*Takifugu rubripes*) and spotted green pufferfish (*Tetraodon nigroviridis*) genomes. 33 markers had unique hits across at least five out of eight species (Appendix: Chapter 3: Additional file 5). *T. nigroviridis* and *O. niloticus* show the highest level of synteny with Atlantic halibut and each other (Appendix: Additional file 5). The order of the markers selected for SNP genotyping in the regions point toward one 3.2 Mb region embedding more than 60 annotated genes. (See direct links to the Ensembl 68 in the Appendix: Additional file 5). No genes associated with sexual differentiation or determination were identified in this region.

3.4 Discussion

Atlantic halibut is a species of increasing commercial interest for cold-water marine aquaculture. However one of the main limitations to profitable culture of the species is the sexual dimorphism in age at maturation related to gender specific growth performance (Bjornsson, 1995; Babiak et al., 2012). To address this bottleneck, the current study demonstrates, for the first time in this species, that indirect monosex

female production is possible for commercial Atlantic halibut aquaculture. While having strong commercial application, this research also had the fundamental aim to investigate the genomic regulation of sex-determination in the species through state-of-the-art high-throughput sequencing methodologies.

RAD-tag sequencing has recently been used with a number of different fish species since the technology became available in 2008. One of the aims of the Baird et al. (2008) study, which first validated the technique in fish, was to fine map QTLs in three-spine stickleback. A number of different restriction-digest methodologies already existed, using high-throughput sequencers. However, what sets the RAD-tag methodology apart is the fact that it combines control over the fragment size that result from the digestion with deep sequencing across individuals, making the identified SNP reproducible (McCormack et al., 2011). This makes the RAD platform very efficient for constructing genetic maps and QTL studies.

In the present study, a genetic map of 5,703 SNPs and 7 microsatellites spanning 1,514 cM was constructed. To our knowledge this is the first dense genetic map incorporating SNPs in any flatfish species. The map has 24 linkage groups, corresponding to the number of chromosome pairs in *H. hippoglossus* (Brown et al., 1997). In a similar study, Amores et al. (2011) constructed a genetic map for spotted gar (*Lepisosteus oculatus*) consisting of 8,406 SNPs. The above map was used to prove that spotted gar diverged from teleosts before the Teleost genome duplication. Genetic maps of more than 4,500 SNPs using RAD-seq were also constructed in rainbow trout (Miller et al., 2012) and in zebrafish (Anderson et al., 2012).

The high LOD (> 10), which was used to assign the genetic markers in linkage groups in our study, ensures that the map is of high quality. However, it must be acknowledged that even though the assignment of markers in linkage groups is robust, none of the available algorithms used for ordering markers provides an accurate positioning of closely spaced markers due to the relatively low number of meiosis represented in our sample size. In a species like Atlantic halibut with no sequenced genome available, a genetic map is an invaluable tool for mapping any trait of interest in a QTL study. Apart from mapping QTL, the identified SNP of the genetic map can be used to construct a genomic relationship matrix, which can replace the relationship matrix inferred by pedigree for calculating breeding values. This would improve accuracy of estimated breeding values (EBV) under Best Linear Unbiased Prediction (BLUP) methodology in a breeding program (Goddard, 2009). The improvement in accuracy is due to the fact that the genomic relationship matrix accounts for the random segregation of chromosome segments at meiosis between siblings.

In this study we associated mapped RAD-tag markers to sex-determination. A major QTL involved in sex-determination was identified in LG 13 in both families (LOD = 12.16 and 6.83 in Family B and C respectively). The location of the above QTL spans a region of around 22 cM. This region should contain one or more genes responsible for sex-determination in Atlantic halibut. The reduced recombination in this region resulted in an almost flat likelihood surface for this region. Genome regions with reduced recombination are a common characteristic of sex chromosomes. In a similar study by Anderson et al. (2012) where the objective was

to identify QTLs involved in sex-determination in zebrafish using RAD-tag, a region in chromosome four spanning more than 20 cM showed reduced recombination. In general suppression of recombination keeps together genes (or alleles) with functions that are advantageous for one sex and avoids their transfer to the other sex chromosome, where they might have negative effects on the opposite sex (Volff et al., 2007).

Our data support the hypothesis that the *H. hippoglossus* has an XX/XY sex determination system. Among flatfish species, Bastard halibut (*Paralichthys olivaceus*) has also been shown to possess an XX/XY system (Yamamoto, 1999), although temperature also influences sex ratio. On the other hand other closely related species, in which sex associated genetic markers have been identified, such as Pacific halibut (*Hippoglossus stenolepis*) (Galindo et al., 2011), spotted halibut (*Verasper variegates*) (Ma et al., 2010), turbot (*Scophthalmus maximus*) (Martinez et al., 2009) and half-smooth tongue sole (*Cynoglossus semilaevis*) (Liao et al., 2009) were all shown to have a ZZ/ZW sex-determination system. Unusually in this group, half-smooth tongue sole has differentiated W and Z chromosomes (Liao et al., 2009). Validating the results of the QTL-Association Analysis is of the utmost importance. The fact that the sex-associated SNPs showed strong association when tested in a wider panel of three families and 36 wild broodstock provides clear evidence that those markers are in strong linkage disequilibrium with the sex-determining gene(s). Marker-assisted selection (MAS) could be conducted using these SNPs, providing a valuable tool towards more efficient production of all-female stocks for the aquaculture industry. In the current study it took four years from initiation of sex

reversal treatment to completion of progeny testing for neomale identification with guaranteed all-female production from the following year. By employing MAS however it would be possible to confirm sex associated genotype from a non-destructive biopsy sample in hormonally-treated fish within 6-12 months of treatment, allowing neomales to be isolated and used from first maturation at three-four years post-treatment. SNPs *Hhi*58665, *Hhi*10170, *Hhi*41238 and *Hhi*47769 are the strongest candidates for MAS since they correctly assign sex in more than 97% of the screened individuals. They span a narrow region of 3.2 cM. Genotyping a larger population for the SNPs in this region would allow fine mapping of the sex-determining locus. Other genetic factors involved in sex determination might also be involved.

The application of this technology will enable the industry to include a greater number of neomales from a wider genetic base to be included in future breeding programmes without the reduction in effective population size (Ne) associated with the use of a small number of neomales from these initial sex-reversed families. Limited examples exist of practical application of MAS in breeding programmes in aquaculture. A Y-specific DNA marker was used to assist in the development of mono-sex female culture in Chinook salmon (*Oncorhynchus tshawytscha*) (Devlin et al., 1991). More recently, MAS has been apply to a QTL for Infectious Pancreatic Necrosis Virus resistance in Atlantic salmon (*Salmo salar*) where initially microsatellite markers were used, and more recently SNPs derived from RAD sequencing have been added (Houston et al., 2008; Houston et al., 2012).

3.5 Conclusions

Overall this work has demonstrated that all-female halibut production is commercially possible using indirect mono-sex production techniques. This in itself confirms that Atlantic halibut has an XX/XY sex-determination system. RAD sequencing produced 90,105 unique loci, and a single sex-determination locus was mapped to LG 13. A further set of 4 markers that were present only or predominantly in DNA from male fish was isolated from two families and validated in a wider population screening, opening the possibility of MAS for sex in the species. Synteny analysis showed that DNA sequences containing Atlantic halibut sex-associated SNPs were consistently clustered in several other genomes, which provides a new focus for research into the sex determination mechanism in this species.

Acknowledgments

We are grateful for support from the Marine Alliance for Science and Technology for Scotland (MASTS), the Scottish Aquaculture Research Forum (SARF 027) and a SPARK award from the Biosciences Knowledge Transfer Network. We thank staff at The GenePool Genomics Facility, especially Urmi Trivedi and Marian Thomson, for assistance with sequencing.

Chapter 4

Nile tilapia

4 Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochomis niloticus*) using RAD sequencing

Christos Palaiokostas [1], Michaël Bekaert [1], Mohd G.Q. Khan [1], John B. Taggart [1], Karim Gharbi [2], Brendan J. McAndrew [1], David J. Penman [1]

[1] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2] The GenePool, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK

Keywords: *Oreochromis niloticus*, Sex determination, QTL Mapping, RAD-seq, Aquaculture.

Abbreviations: RAD: restriction-site associated DNA; SNP: single nucleotide polymorphism; QTL: Quantitative trait locus; LG: Linkage group.

Status: *PLoS ONE 8(7): e68389. doi:10.1371/journal.pone.0068389*

Contributions: The first draft of the present manuscript was compiled and written in full by the author of this thesis, who was also fully involved in all subsequent revisions. Fish breeding, rearing, sexing (fish used in KASP assays genotyping), DNA extraction, preparation of the RAD libraries (under guidance of John Taggart), genotyping for KASP assays, construction of the genetic map, QTL mapping and association analysis were conducted by the author of this thesis. The other co-authors contributed towards the experimental design, sexing and breeding the fish for the RAD libraries and in the analysis of the sequenced reads.

Abstract

Sex in Nile tilapia (*Oreochromis niloticus*) is principally determined by an XX/XY locus but other genetic and environmental factors also influence sex ratio. Restriction-site Associated DNA (RAD) sequencing was used in two families derived from crossing XY males with females from an isogenic clonal line, in order to identify Single Nucleotide Polymorphisms (SNPs) and map the sex-determining region(s). We constructed a linkage map with 3,802 polymorphic SNP markers, which corresponded to 1,646 discrete map positions, and identified a major sex-determining region on linkage group 1, in the same location for both families, explaining nearly 96% of the phenotypic variance. This sex-determining region was mapped in a 2 cM interval, corresponding to approximately 1.2 Mb in the *O. niloticus* draft genome. In order to validate this, a diverse family (4 families; 96 individuals in total) and population (40 broodstock individuals) test panel were genotyped for five of the SNPs showing the highest association with phenotypic sex. From the expanded data set, SNPs *Oni*23063 and *Oni*28137 showed the highest association, which persisted both in the case of family and population data. Across the entire dataset all females were found to be homozygous for these two SNPs. Males were heterozygous, with the exception of five individuals in the population and two in the family dataset. These fish possessed the homozygous genotype expected of females. Progeny sex ratios (over 95% females) from two of the males with the "female" genotype indicated that they were neomales (XX males). Sex reversal induced by elevated temperature during sexual differentiation also resulted in phenotypic males with the "female" genotype. This study narrows down the region containing the main sex-determining locus, and provides genetic markers

tightly linked to this locus, with an association that persisted across the population. These markers will be of use in refining the production of genetically male *O. niloticus* for aquaculture.

4.1 Introduction

Patterns of sex determination and differentiation in fish are very varied, with a wide range of gonochoristic and hermaphroditic species. Among the gonochoristic species, genetic and/or environmental factors determine sex (Devlin & Nagahama, 2002; Penman & Piferrer, 2008; Kamiya et al., 2012). The first sex-determining gene isolated in a fish species was *DMY/dmrt1bY* in medaka (*Oryzias latipes)* (Matsuda et al., 2002; Yano et al., 2012).

More recently, several other fish sex determining genes have been isolated: *Gsdf(Y)* in Luzon ricefish (*Oryzias luzonensis*) (Nanda et al 2003, Myosho et al 2012); *amhy* in Patagonian pejerrey (*Odontesthes hatchery*) (Devlin & Nagahama, 2002; Hattori et al., 2012); *Amhr2* in tiger pufferfish (*Takifugu rubripes*) (Kamiya et al., 2012); and *sdY* in rainbow trout (*Oncorhynchus mykiss*) (Yano et al., 2012a). Four of these genes come from those involved in sexual differentiation, while one derives from an immune-related gene (the *sdY* protein in rainbow trout is similar to part of the interferon regulatory factor 9). In all of these species the sex determination system is principally XX/XY, but mismatches between sexual genotype and phenotype are not uncommon (Nanda et al., 2003), sex reversal can be induced, YY males are viable in several species, and differentiated sex chromosomes are relatively uncommon (Devlin & Nagahama 2002). The consequences of such variation are that sex

determination genes in fish may likely require identification at the species or genus level. Furthermore, while we might search for genes from the sex differentiation pathway in a region where a sex determination gene has been mapped, it is by no means certain that we can easily identify a master switch gene.

Since many farmed species of fish exhibit sexual dimorphism in a range of traits of interest like growth or age at maturity, clarification of the sex-determining system of such fish is beneficial for the aquaculture industry towards the production of mono-sex stocks. Nile tilapia (*Oreochromis niloticus*) is one of the most important farmed species with a production exceeding 2.8 million metric tonnes in 2010 (FAO 2012). Intensive commercial production generally requires all-male stocks, not only because males grow faster but also to avoid uncontrolled reproduction before harvest.

In tilapias, evidence so far suggests the existence of two different major sex-determining systems. In some tilapia species, including Nile tilapia and Mozambique tilapia (*Oreochromis mossambicus*), sex is primarily determined by an XX/XY system on linkage group (LG) 1, whereas in others, for example blue tilapia (*Oreochromis aureus*), sex is primarily determined by a WZ/ZZ system on LG 3 (Cnaani et al., 2008). However, other factors may influence sex determination and differentiation. In Nile tilapia, genes on LG 3 (Karayücel et al., 2004) and LG 23 (Eshel et al., 2010; Eshel et al., 2012), and temperature (Baroiller et al., 2009) can affect sex ratio. Crosses between YY males and XX females generally give less than to 100% male progeny predicted from a simple XX/XY system (Beardmore et al., 2001). Many of the studies on sex determination in Nile tilapia have been carried out

on fish derived from Lake Manzala in Egypt, the subject of the present study, and it is clear that both non-LG 1 genes and temperature affect sex ratios in at least some families in this population.

The current linkage map for Nile tilapia is based on more than 500 markers, mostly microsatellite (Lee et al., 2005). The sex-determining region has been previously mapped close to microsatellite *UNH995* on LG 1 (Lee et al., 2003). This region contains two genes implicated in vertebrate sexual differentiation, wt*1b* and *cyp19a*, but further mapping ruled these out as candidates for the major sex-determining locus (Lee & Kocher, 2007). Restriction site associated DNA (RAD) sequencing (Baird et al., 2008) offers the possibility to construct much higher density linkage maps in a cost-efficient manner. In this study we used RAD sequencing to identify single nucleotide polymorphisms (SNPs) in two crosses between XY males and females from an isogenic clonal line. A genetic map was constructed based on 3,802 SNP markers. A quantitative trait locus (QTL) analysis was conducted based on these SNPs and was followed by an association analysis for the SNPs that showed the highest association with phenotypic sex using a diverse dataset of both family and population data. Altogether these data located the sex-determining QTL in a region of approximately 1.2 Mb on LG 1.

4.2 Materials and Methods

4.2.1 Sample collection and preparation

The fish used in this study came from the Tropical Aquarium Facilities of the Institute of Aquaculture at the University of Stirling. They originated from a population that was established in 1979 from Lake Manzala, Egypt (31°16'N, 32°12'E). All working procedures complied with the Animals Scientific Procedures (Parliament of the United Kingdom 1986 Animals Act 1986). Fish were reared in recirculating water systems at 27-28°C, and fed on commercial trout diet (Trouw Aquaculture Nutrition, UK; manufacturer Skretting, UK). To set up the families used in this study, mature females were held in glass aquaria and eggs were manually stripped following ovulation. Milt was manually stripped from male fish and used to fertilise the eggs *in vitro*. Eggs were incubated in downwelling incubators until the larvae had absorbed the yolk sac. Fry from families 1-6 (Table 1) were then transferred to tanks in recirculating systems and reared for 3-4 months before being killed and sexed by microscopic examination of the gonads (Guerrero & Shelton, 1974). A sample of fin tissue was taken and fixed in 100% ethanol for DNA extraction. Family 7 was split at yolk sac absorption: one group of 80 fry was reared at 36°C for ten days (Wessels & Hörstgen-Schwark 2007) in a static 5 L tank, while a control group (80 fry) was reared at 28°C. The survival of the two groups was 88% and 91% respectively. Subsequent rearing and sexing was as for families 1-6.

Families 1 and 2 (68 offspring and 20 offspring respectively) with dams from an isogenic XX clonal line and XY sires (as judged from balanced sex ratios in crosses to clonal line and outbred females and a high association between phenotypic sex and the LG 1 *UNH995* marker) were used to prepare RAD Libraries (Table 4.1). The available genome draft of Nile tilapia is based on females from this clonal line. The sex associated SNPs were further validated by genotyping four further families with balanced sex ratios (Families 3-6: 24 offspring each) and broodstock (40 individuals; Table 4.1). These SNPs were finally used for genotyping a family in which elevated temperature induced a change in sex ratio (Family 7) to test whether the above SNPs could be useful in distinguishing neomales (XX males) from normal males. The sex ratio of the control group (reared at 28°C) did not show any deviations from the expected 1:1 ratio, while in the high temperature (36°C) treated group the proportion of males exceeded 96%.

Table 4-1. Fish used in the study-background information.

ID	Use	Sire strain	Dam strain	No. of females	No. of males	Total
Family1	RAD Libraries	Red†	Clonal	33	35	68
Family2	RAD Libraries	Red†	Clonal	10	10	20
Family3	SNP Assays	Red†	Wild*	12	12	24
Family4	SNP Assays	Red†	Wild*	12	12	24
Family5	SNP Assays	Wild*	Clonal	12	12	24
Family6	SNP Assays	Red†	Clonal	12	12	24
Family7 (28°C)	SNP Assays	Red†	Wild*	34	33	67
Family7 (36°C)	SNP Assays	Red†	Wild*	4	66	70
Clonal line	SNP Assays	-	-	0	2	2
Red strain	SNP Assays	-	-	6	15	19
Wild strain	SNP Assays	-	-	12	5	17

*"wild" refers to wild type coloration; † "red" refers to red body colour, which is controlled by a single gene

4.2.2 RAD library preparation and sequencing

DNA was extracted from fin tissue of the fish using the REALPure genomic DNA extraction kit (Durviz S.L.) and treated with RNase to remove residual RNA from the sample. Each sample was quantified by spectrophotometry (Nanodrop) and quality assessed by agarose gel electrophoresis, and was finally diluted to a concentration of 50 ng/μL in 5 mmol/L Tris, pH 8.5. The RAD library preparation protocol followed essentially the methodology originally described in Baird et al. (2008) and comprehensively detailed in Etter et al. (2011), with the minor modifications described in Houston et al (2012). The RAD specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. (2011).

Each sample (0.72 μg parental DNA / 0.24 μg offspring DNA) was digested at 37°C for 40 minutes with *Sbf*I (recognising the CCTGCA|GG motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U *Sbf*I per μg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 μg DNA per 50 μL reaction volume. The reactions (12 μL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual specific P1 adapters, each with a unique 5 bp barcode (Appendix, Table S1), were ligated to the *Sbf*I digested DNA at 22ºC for 60 minutes by adding 1.8 / 0.6 μL 100 nmol/L P1 adapter, 0.45 / 0.15 μL 100 mmol/L rATP (Promega), 0.75 / 0.25 μL 10× Reaction Buffer 2 (NEB), 0.36 / 0.12 μL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 45 / 15 μL with nuclease-free water for each parental / offspring sample. Following heat

inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in appropriate multiplex pools (Table S1). Shearing (Covaris S2 sonication) and initial size selection (250-550 bp) by agarose gel separation was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation, library amplification, exactly as in the original RAD protocol (Baird et al., 2008; Etter et al., 2011). A total of 150 μL of each amplified library (16-18 PCR cycles, library dependent) was size selected (c. 350-650 bp) by gel electrophoresis. Following a final gel elution step into 20 μL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in two lanes of an Illumina HiSeq 2000, one run using 100 base paired-end reads, the other 100 base single reads (v3 chemistry). Raw reads were process using RTA 1.12.4.2 and Casava 1.6 (Illumina). The reads were deposited at the NCBI BioProject under the accession SRP017804.

4.2.3 Genotyping RAD alleles

Reads of low quality (score under 30, while the average quality score was 37), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotyped using Stacks software 0.9995 (Catchen et al., 2011). The likelihood-based SNP calling algorithm (Hohenlohe et al 2010) implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. A minimum stack depth of at least 30 and a maximum of 2 mismatches were allowed in a locus in

an individual and up to 1 mismatch between alleles. The pair-ends were assembled using Stacks and Velvet version 1.2.08 (Zerbino & Birney, 2008) and used to separate RAD-tag sequence with or without potential SNP but belonging to separate loci (duplication products). Polymorphic RAD-tags may contain more than one SNP, but the vast majority (over 99%) showed only two allelic versions.

4.2.4 Genetic map construction

The genetic map was constructed using R/Onemap (Margarido et al., 2007) and TMAP (Cartwright et al., 2007). The allocation of markers into linkage groups was conducted using R/Onemap. This package uses Hidden Markov Models (HMM) algorithms for outbred species while in parallel implements the methodology described in Wu et al. (2002) for calculating the most probable linkage phase. Linkage groups were formed using a minimum LOD value of 6. TMAP was used to order the markers in every linkage group. By using an HMM maximum likelihood model and taking into account potential genotypic errors it reduces the tendency to erroneously derive oversized linkage groups, a phenomenon which is often observed in dense maps. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function (Appendix: Table S2). The linkage group name (number) was subsequently matched with the Broad Institute of MIT and Harvard genome assembly Orenil1.1 (NCBI Assembly GCA_000188235.2). The genetic map was drawn and aligned using Genetic-Mapper v0.3 (Bekaert, 2012).

4.2.5 QTL mapping

The QTL analysis was performed using R/qtl (Broman & Sen, 2009). With the dam originating from a clonal line and by inferring the most probable phase of the genetic markers of the sire the cross had the same properties as a backcross and was analysed as such. Initially existence of single QTLs was tested (R/qtl function: *scanone*). The model used for the analysis was based on interval mapping. The algorithm used considers the phenotype to follow a mixture of Bernoulli distributions and uses a form of the expectation-maximisation algorithm for obtaining maximum likelihood estimates (Broman & Sen, 2009). Permutation tests (10,000 permutations) were conducted in order to correct for the multiple testing. A multidimensional approach towards QTL mapping was adopted by using models from R/qtl that accounted for the QTL in LG 1, for two-QTLs and for multiple-QTLs simultaneously (R/qtl functions: *makeqtl, addqtl*, *scantwo, fitqtl, stepwiseqtl*). With this approach greater power can be achieved in the analysis allowing for detection of QTLs that would be remained undetected in the one-dimensional approach above.

Approximate Bayesian 95% density intervals were calculated. An approximate estimate of the phenotypic variance explained by the QTL is obtained from the following equation: $1\text{-}10^{-2LOD/n}$. It has to be stressed that while the estimated variance may be reasonable for additive QTL, problems can be caused in the case of linked QTL (Broman & Sen, 2009).

4.2.6 SNP assays

We designed SNP assays using the KASP genotyping system (KBioscience UK Ltd) for five SNPs (*Oni*20117, *Oni*61067, *Oni*23063, *Oni*28137, *Oni*22734, NCBI dbSNP accession 748775078, 748775079, 748775085, 748775081 and 748775082 respectively; Appendix: Table S3) that showed the highest association with sex in the two families that were used for the RAD-seq (Table S4). Allele-specific primers and other assay components were supplied by KBioscience UK Ltd, based on the supplied marker sequences (Appendix: Table S3, Data S1). PCR reactions were carried for 10 μL final volume reactions. The cycling conditions were the following: 94ºC for 15 min, 94ºC for 20 sec, touchdown over 61ºC to 55ºC for 60 sec (10 cycles dropping 0.8ºC each cycle) and an extra 34 cycles at 55ºC.

4.2.7 Association analysis in family and population data

Family data (offspring from four different families; 96 individuals in total) and 40 unrelated broodstocks (Table 4.1) were genotyped for the above five SNP markers. A family showing high response in elevated temperature was also genotyped for the same SNPs in order to check whether those SNPs could be used in distinguishing neomales from normal males. An association analysis was performed using R/SNPassoc (González et al., 2007). In the case of family data, association was tested both in separate families and across all families together. A Bernoulli generalised linear model was applied in order to test the magnitude of association between the SNP genotypes and phenotypic sex using this package (function

association). Both the Bonferroni and permutation tests (10,000 permutations) were used in order to correct for multiple testing.

4.3 Results

4.3.1 RAD sequencing

Two crosses with 68 and 20 offspring and their parents (including clonal females) were sampled (Appendix: Table S1). The samples were barcoded, pooled and sequenced in two lanes of an Illuminia HiSeq 2000 sequencer. In total, 459,543,173 raw reads (101 bases long) were produced (171,201,863 paired-end and 117,139,447 single-end reads, NCBI BioProject SRP017804). After removing low quality sequences (quality score under 30), ambiguous barcodes and orphaned paired-end reads, 87.03% of the raw reads were retained (399,918,024 reads). The Stacks package (Catchen et al., 2011) was then used to make the assembly of the sampled loci from each individual: 95,791 and 127,014 RAD-tags (or RAD markers) were retrieved for Families 1 and 2 respectively, covering 152,916 RAD-tags in total including 69,889 of these shared between the two families (Figure 4.1). The number of reads and RAD-tags for each sample are reported in the Appendix.

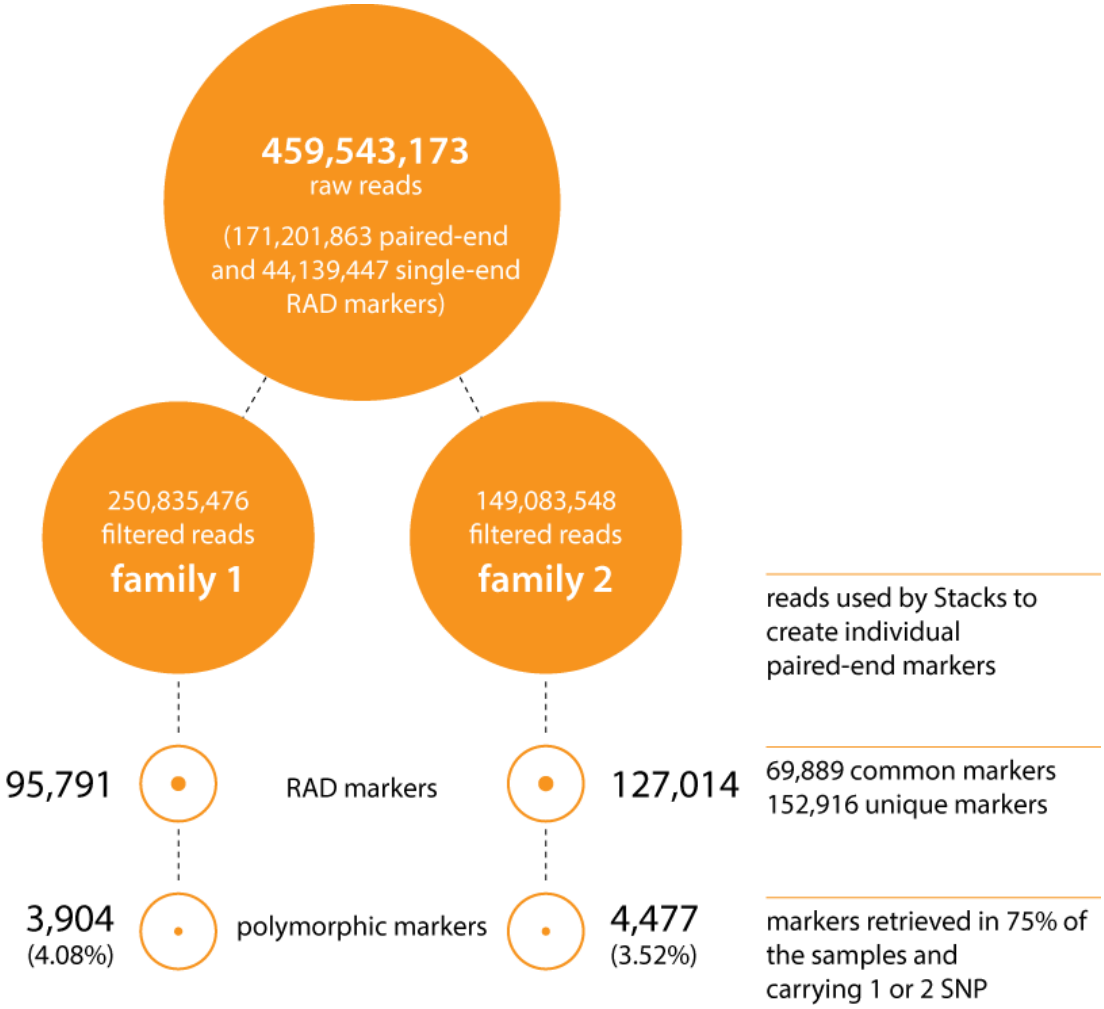

Figure 4-1. Sequencing and RAD-tag summary. Details of the number of reads before and after filters (orange disk) followed by the reconstructed number of RAD markers and polymorphic RAD markers (orange circles).

4.3.2 Genetic map

In order to maximise the number of informative markers and minimise the amount of missing or erroneous data, we used SNP markers retrieved in at least 75% of the samples in each family, and carrying one or two SNPs. Since Family 2 had only 20 offspring, the genetic map was constructed with the Family 1 data only (68 offspring), while Family 2 was used to validate this. The sire-based map consists of 3,802 SNP markers, with 1,646 discrete map positions, that were grouped in 23

linkage groups, with an average spacing of 0.7 cM and spanning a total distance of 1,176 cM (Figure 4.2 and Tables 4.2 & Appendix:Table S2). In the second family (Family 2, 20 offspring) 724 of the above markers were heterozygous. The linkage groups were named according to the Broad Institute of MIT and Harvard genome assembly Orenil1.1 (NCBI Assembly GCA_000188235.2). We were not able to join the markers into 22 linkage groups corresponding to the 22 chromosomes expected from the karyotype (Majumdar & McAndrew, 1986). However, by comparing our map to the draft tilapia genome sequence, two linkage groups were coalesced to form LG 3 in Figure 4.2. LG 3 contains a broad region of recombination suppression (Cnaani & Hulata, 2008).

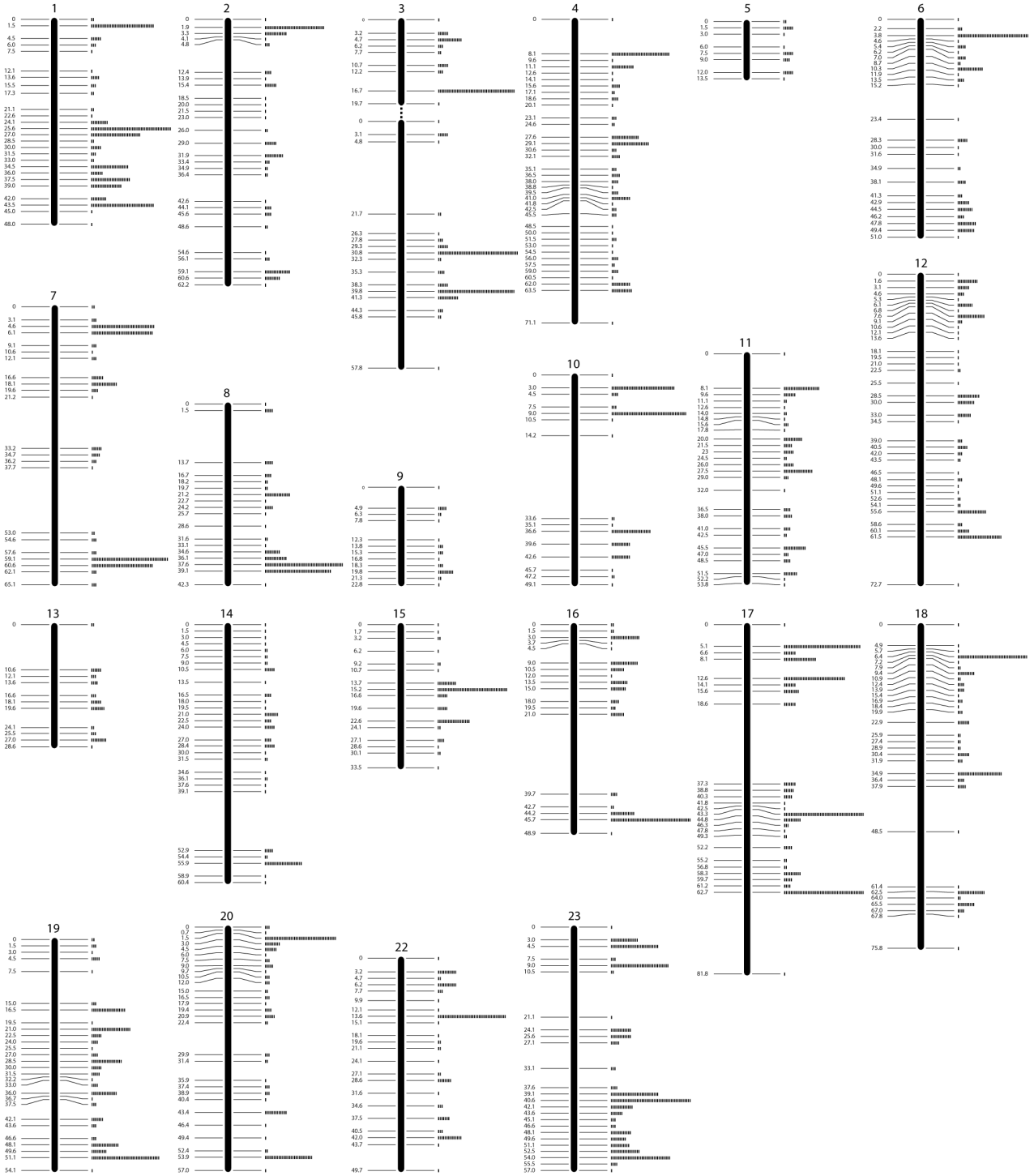

Figure 4-2. Nile tilapia linkage map. Map with linkage group assignment determined using syntenic markers with previously published *O. niloticus* maps. The positions on the left side of the chromosomes are in cM. The rectangles on the right hand side represent the number of markers at this position. The numbering of the linkage groups corresponds to that in the Broad Institute genome anchored assembly Orenil1.1

Table 4-2. *O. niloticus* genetic map based on sire only information.

Linkage Group	No. of markers	No. of discrete map positions	Length (cM)
1	284	115	48.0
2	136	63	62.2
3a	109	40	19.7
3b	153	64	57.8
4	191	112	71.1
5	27	15	13.5
6	161	65	51.0
7	232	81	65.2
8	170	76	42.3
9	32	23	22.8
10	143	43	49.1
11	135	80	53.8
12	173	94	72.7
13	47	26	28.6
14	90	59	60.4
15	192	66	33.5
16	154	61	48.9
17	323	121	81.5
18	188	80	75.8
19	246	106	54.1
20	155	77	57.0
22	117	60	49.7
23	344	119	57.0
Total	**3,802**	**1,646**	**1,176**

4.3.3 QTL mapping

The results from the single-QTL model for binary traits provided evidence for the existence of a major QTL in LG 1 for both families (Figure 4.3A). The highest logarithm of odds (LOD) score for Families 1 and 2 were 18.50 and 6.02 respectively and the QTL was observed in the same location for both families with the same SNPs showing significant linkage (Figure 4.3B). The difference in the LOD scores is due to the smaller number of meioses in the second family (20 offspring). The fact that the marker phase was the same in the two families (in the QTL region) allowed in addition to the above a joint analysis of the two families. The genome wide

threshold LOD value ($\alpha = 0.001$) was 3.89 as calculated from permutation tests (10,000 permutations). The highest LOD score was observed at the 28.5 cM of LG 1. Two adjacent SNPs were located in the above position (*Oni*23063, *Oni*68581, from paired RAD-tags on either side of a single *Sbf*I restriction site). Only *Oni*23063 was analysed subsequently.

The two-QTL model did not reveal any significant additional QTL or any evidence for epistasis. Even though models that take into account the existence of a major QTL (as in this study) or ones that test for existence of multiple QTLs simultaneously reduce the residual variation (providing higher power in the analysis for detecting additional QTLs at least of modest effect), no additional QTLs were detected. The calculated 95% Bayesian Density Intervals for the QTL location spanned a region of 2 cM (28-30 cM in LG 1), corresponding to approximately 1.2 Mb in the *O. niloticus* genome. The phenotypic variance explained by the QTL was 96%.

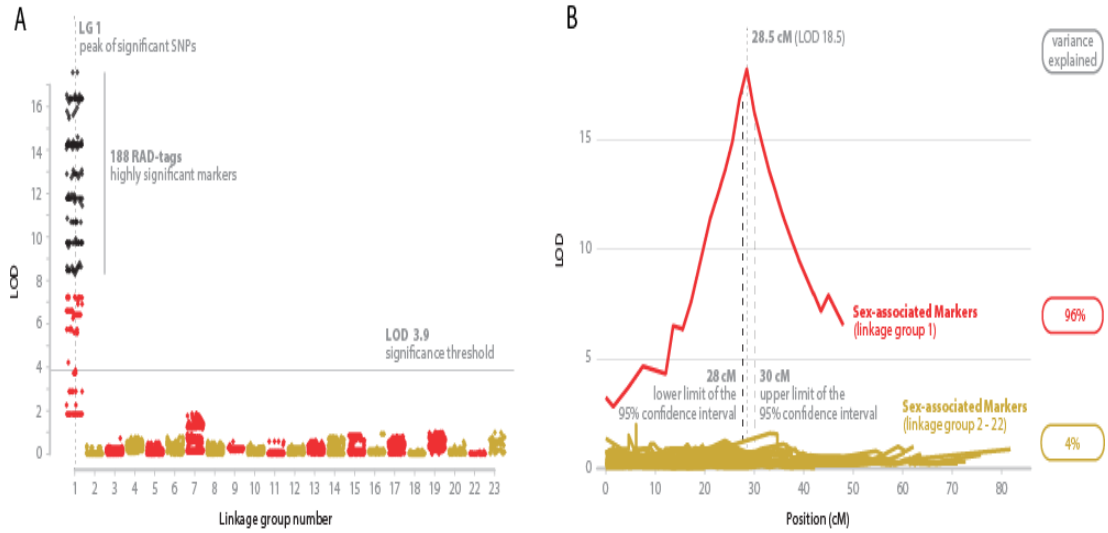

Figure 4-3. Results from QTL-Analysis in Nile tilapia offspring. (A) Association results for genotyped SNPs. SNPs with *P*-values achieving genome-wide significance ($p<7.2\times10^{-8}$) are shown in black. (B) Regional analysis of the QTL. Plot of the LOD score (sex-association QTL search) along the linkage groups

4.3.4 Association analysis using family and population samples

The five SNP markers that showed the strongest linkage with sex (Figure 4.4A) in the two mapping families were tested in a larger panel consisting of family and population samples. The *p*-value thresholds ($\alpha = 0.05$, multiple test correction) for permutation and the Bonferroni correction tests were 0.016 and 0.01 respectively. All five SNP markers were found to be significantly associated with sex in the family data both when tested in each family separately and across all families simultaneously (Figure 4.4A). For the population data (40 broodstock), four of the five SNP markers showed significant association with phenotypic sex (the exception was *Oni*20117, $p = 0.73$). *Oni*23063 and *Oni*28137 showed the highest association with phenotypic sex for both the population and family data (*p*-values of 1.08×10^{-7} and 3.024×10^{-29} respectively). Females were homozygous and males heterozygous for those two SNPs. The only exceptions from the above pattern were found for one

male progeny from the four tested families and five male broodstock fish (Appendix: Chapter 4: Table S4). Progeny testing of two of these five male broodstock in crosses to XX females) gave nearly all-female progeny (over 95% females). No recombination was observed for those two SNPs in the entire dataset that was genotyped.

The sex-determining region location spanned a distance of 2 cM, corresponding to approximately 1.2 Mb in the *O. niloticus* genome, with the peak of both LOD and association located at the position of *Oni*23063. The two markers most strongly associated with sex, *Oni*28137 and *Oni*23063, are 400 kb apart in the *O. niloticus* genome. The 10 annotated genes of this region of the genome are therefore potential candidates (Figure 4.4B and Appendix: Chapter 4: Data S2).

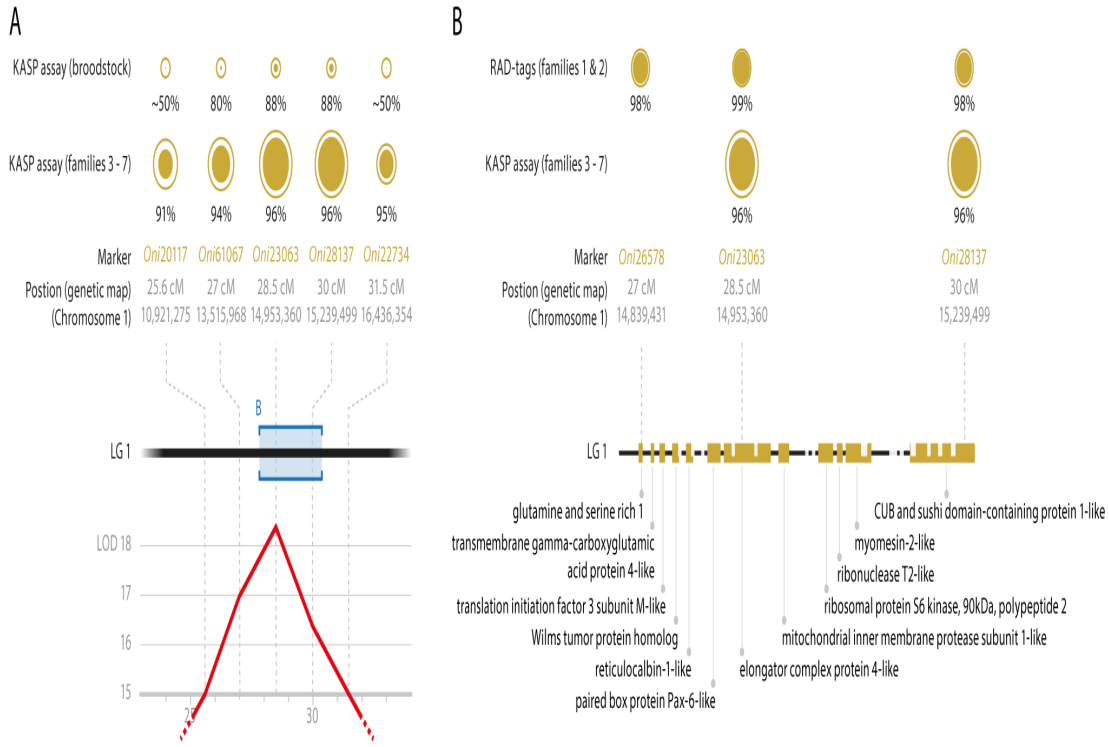

Figure 4-4. KASP assay and fine gene mapping on LG 1. (A) Details of the five markers tested by KASP assay. From bottom to top: LOD score (QTL sex-association); Location of the five markers (in the genetic map in cM and anchored draft genome in bp); KASP assay results. The outer circle diameters are proportional to the number of alleles tested. The inner (solid) disks represent the marker association with the phenotypic sex. Detailed data are provided in Appendix (Chapter 4: Table S4). (B) Details of the region of higher association. The bottom half is a schematic (to scale) of the chromosome LG 1, scaffold 17. It includes 14 gaps (white gaps) in the genome and 10 annotated genes (orange boxes). The sex-determining factor is located between "glutamine and serine rich 1" and "CUB and sushi domain-containing protein 1- like"

4.3.5 Sex reversal

To explore the effects of sex reversal on the association between genotype and phenotypic sex, since family specific QTLs involved in temperature sex-determination have been recently identified (Lühmann et al., 2012), a family that showed a significant effect of raised temperature on sex ratio was genotyped for the two SNPs showing the strongest association (*Oni*23063 and *Oni*28137). While these showed highly significant association ($p = 4.05 \times 10^{-14}$) in the control group, in the

temperature treated group neither marker was significantly associated after correction for multiple testing ($p = 0.039$). Five phenotypic males were found to deviate from the expected association in the control group, while in the treated group 28 phenotypic males deviated from the expected association for *Oni*23063 and *Oni*28137 (Figure 4.5).

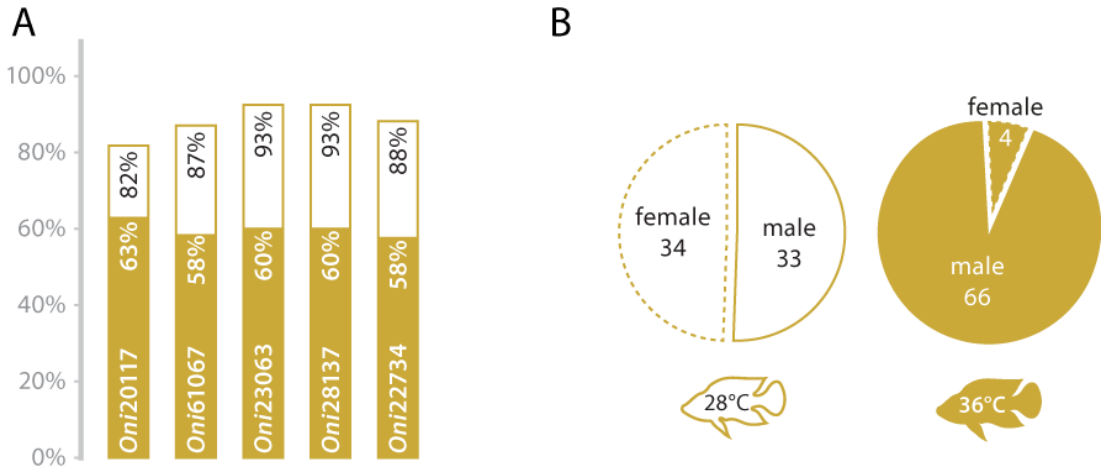

Figure 4-5. Sex reversion tests (family 7). (A) KASP assay results in family 7 offspring in response to elevated temperature (36°C) compare to standard 28°C. Each bar represent the ratio of male/female at 36°C (solid orange) and at 28°C (white background). (B) Detail of the phenotypic sex in response to elevated temperature (male have continuous border, female are presented with a discontinuous border).

4.4 Discussion

Current evidence suggests that Nile tilapia possess an XY/XX male heterogametic system complicated by genetic variance at this and other loci, environmental factors and probably by their interaction (Cnaani & Hulata, 2008). While previous family-based studies provided evidence for the existence of a major sex determining region on LG 1 in Nile tilapia (Cnaani et al., 2008; Lee et al., 2003; Lee & Kocher, 2007; Lee et al., 2011), various anomalies, including inconsistencies among families, have been observed. For example, Lee et al. (2003) demonstrated the existence of a sex

determining QTL on LG 1, using microsatellite markers, which was mapped to an interval of 10 cM. However, the association between this region and phenotypic sex was only observed in two out of the three crosses studied. In the third cross, no association was observed with any genomic region. Frequent departures have been observed from the 100% male progeny in YY × XX crosses predicted from a single XX/XY locus (Beardmore et al., 2001; Mair et al., 1997). Eshel et al. (2010, 2012) found that markers in LG 23 showed the highest association with phenotypic sex in a cross in a population of *O. niloticus* also derived from Lake Manzala in Egypt. However no information was provided whether the above QTL on LG 23 persists in other crosses as well.

In the present study, our strategy was to use RAD sequencing to develop a much higher density linkage map than that of Lee et al. (2005), based on carefully selected mapping families (isogenic female crossed to normal male; balanced sex ratio; progeny already tested for association between LG 1 microsatellite markers and phenotypic sex), then validate a set of tightly sex-linked markers in further family and population samples to test for population-wide association.

SNPs *Oni*61067, *Oni*23063 and *Oni*28137 showed the highest and most consistent association with phenotypic sex in all our data. SNP *Oni*23063 was the marker with the highest score in the two families that were used for the QTL analysis, while the additional families that were genotyped had the same magnitude of association with *Oni*28137, with no recombination observed between these two markers. *Oni*61067 gave the third largest association. The above provides evidence that the most

probable location for the master sex determining gene is in this region on LG 1, which spans around 1.2 Mb in scaffold 17 in the *O. niloticus* draft genome (Figure 4).

In all tested families, at each of the two loci (*Oni*23063 and *Oni*28137) females were homozygous for the same allele, while males were heterozygous, apart from two males that were homozygous for the allele found in females. The results from the unrelated adult individuals showed again the same pattern, with five males that were homozygous for these two SNPs. The highly significant population-based linkage disequilibrium observed strongly supports the view that these two SNP markers are in very close proximity to the causative locus.

Interestingly, no phenotypic female was mis-assigned in our entire dataset. All females were found to be homozygous for *Oni*23063 and *Oni*28137. Our data support the hypothesis that the mis-assigned males were in fact genetically female fish (as defined by the LG1 sex-determining locus) that have undergone sex reversal due unknown genetic or environmental factor(s). Available progeny data from 2 of these male broodstock indicated that those were in fact sex-reversed males (over 95% females in the progeny data).

When a family responsive to elevated temperature was genotyped for *Oni*23063 and *Oni*28137, the temperature treated group showed a significant number of phenotypic males with the female compatible genotype, while the typically strong association between SNP genotype and phenotypic sex was found in the control group. Rearing

at high temperatures (36°C) results in significant masculinisation in some progenies if started around 10 days post fertilisation and applied for at least 10 days (Baroiller et al., 1995; Baroiller et al. 2001; Tessema et al. 2006). Wessels & Hörstgen-Schwark (2007) showed that temperature-dependent sex ratio is a heritable trait. Lühmann et al. (2012) mapped family-specific QTL for high temperature masculinisation effects to LG 1, LG 3 and LG 23. There is also evidence for feminisation by high temperature, from crosses expected to produce all-male progeny (Abucay et al., 1999; Kwon et al., 2002; Wessels & Hörstgen-Schwark, 2011).

The genomic region, which was derived from the calculated density interval, includes 10 different genes. Even though none of those genes have been known to be implicated in the sex determination pathway of any organism, it would not seem prudent to exclude this possibility, considering the diversity of sex-determining systems in fish. The most striking example was observed in rainbow trout, where a gene associated with immune related functions in other organisms, was found to be its master sex-determining gene (Yano et al., 2012a). However, despite the fact that the results of this study showed that the major sex-determining factor is in close proximity with the tested SNPs, we cannot be confident that the major sex-determining gene is necessarily one of these ten genes. The fact that the available genome draft originated from a female *O. niloticus* means that in case the major sex-determining gene is observed only in the male (as found in the *O. latipes* by Matsuda et al., 2002), no information can be inferred by using the available genome sequence.

Finally there is also the possibility that no annotation about the major sex-determining gene is available in the current genome draft.

This study did not address the reasons why YY × XX crosses yield varying proportions of females. However, the SNP markers developed will help to tackle this issue, a key constraint on the commercial development of this technique.

4.5 Conclusions

This study provides a linkage map for Nile tilapia genome that is several-fold denser than the existing one, a reduced candidate region for the sex determining gene(s) and a set of tightly sex-linked SNP markers. Although we could not identify the causative gene(s), the fact that no female was mis-assigned using our sex-associated SNPs means that those SNPs could be also of high practical value towards the production of all male stocks for the Nile tilapia aquaculture industry.

Acknowledgments

We thank staff at The GenePool Genomics Facility, especially Urmi Trivedi and Marian Thomson, for assistance with RAD library preparation and sequencing.

Chapter 5

Nile tilapia skewed sex ratio study

5 Identification of QTL involved in sex reversal in Nile tilapia (*Oreochromis niloticus*) families with skewed sex ratios using ddRAD-seq

Christos Palaiokostas [1], Michaël Bekaert [1], Mohd G.Q. Khan [1], John B. Taggart [1], Karim Gharbi [2], Brendan J. McAndrew [1], David J. Penman [1]

[1] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2] The GenePool, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK

[#] Corresponding author: Tel +44 (0)1786 467901; Fax +44 (0)1786 472133; Email

Keywords: *Oreochromis niloticus*, Sex determination, QTL Mapping, RAD-seq, ddRAD-seq, Aquaculture.

Abbreviations: RAD: restriction-site associated DNA; SNP: single nucleotide polymorphism; QTL: Quantitative trait locus; LG: Linkage group.

Status: To be submitted

Contributions: The first draft of the present manuscript was compiled and written in full by the author of this thesis, who was also fully involved in all subsequent revisions. Fish breeding, rearing, sexing (fish used in temperature treatment), DNA extraction, preparation of the ddRAD libraries (under guidance of John Taggart), genotyping for KASP assays, construction of the genetic map, QTL mapping and association analysis were conducted by the candidate. The other co-authors contributed towards the experimental design, sexing, breeding the rest of the fish and in the analysis of the sequenced reads.

Abstract

Genomic regions in LG 1, 3 and 23 have previously been shown to be involved in the sex-determining system of Nile tilapia (*Oreochromis niloticus*), probably interacting in certain cases with elevated early rearing temperature as well. This study focused on mapping sex-determining QTL in families with skewed sex ratios. These included one family (family 1) that showed an excess of males when reared at standard temperature (28^{o}C) and another family (family 2) in which an excess of males was observed when fry were reared at 36^{o}C for ten days from first feeding. All the samples used in the current study were genotyped for two SNPs located in the expected major sex-determining region in LG 1 (*Oni23063, Oni28137*). Interestingly the only misassigned individuals were males appearing with the female expected genotype, suggesting that those offspring had undergone sex reversal with respect to the major sex-determining locus. We applied ddRAD-seq to offspring from these two families (family 1:98 offspring; family 2: 120 offspring) using two sequencing lanes of an Illumina HiSeq 2000. Two genetic maps were constructed consisting on 641 and 1155 SNPs from families 1 and 2 respectively. QTL analysis provided evidence for a genome-wide significant QTL in LG 20 in family 1. A significant QTL at the chromosome-wide was also found in the same linkage group in family 2. Evidence was found for an additional sex-determining QTL in family 2, at the beginning of LG 1 in proximity (3.4 Mb) to the aromatase gene *cyp19a1*. Overall the results from this study suggest that these previously undetected QTLs are involved in sex determination in the Nile tilapia, causing sex reversal (masculinization) with respect to the XX genotype at the major sex-determining locus in LG1.

5.1 Introduction

In most fish species, the sex chromosomes are still in early stages of differentiation compared to mammals, and do not show distinct differences in length or gene content (Devlin & Nagahama, 2002). Instances of both the XY/XX male heterogametic and the ZZ/ZW female heterogametic sex-determining systems can be found in fish, while the fact that spontaneous sex-reversed XX males are generally fully fertile could indicate that sex-determining regions could be located also on autosomes. Even though the YY genotype is not compatible with life in mammals, YY and WW genotypes are viable in most fish species indicating that the gene content of the Y and W chromosomes are very similar to that of their X and Z counterparts respectively (Volff et al., 2007; Piferrer & Guiguen, 2008). Generally an even sex ratio can be expected in species with genetic sex determination. However, departures from an equal sex ratio have been observed both in population and family studies, stressing the complexities of sex-determination in fish. Distorted population or family sex ratios are most likely due to environmental sex reversal (Pompini et al 2013), hormonal treatment or interaction between genetic and environmental factors (Penman & Piferrer, 2008). Environmental effects on sex ratios may vary even among different strains of the same species (Mylonas et al., 2005).

Many farmed species of fish exhibit sexual dimorphism in a range of traits of interest like growth or age at maturity, stimulating research to clarify the sex determining system of such fish with the objective of the production of mono-sex stocks for the aquaculture industry. *Oreochromis niloticus* (Nile tilapia) is one of the most

important farmed species with a production exceeding 2.8 million metric tonnes in 2010 (FAO 2012). Intensive commercial production generally requires all-male stocks, not only because males grow faster but also to avoid uncontrolled reproduction before harvest.

Current evidence suggests that O. *niloticus* possesses a complex sex determination system comprising an XY/XX male heterogametic system, other genetic and environmental factors (with temperature being the most important). The major sex-determining region has been previously located on LG 1 (Lee et al., 2003) and fine mapped in a region of approximately 1.2 Mb (Palaiokostas et al., 2013).

Temperature can affect sex ratios in tilapias, with rearing temperatures above 34 ^{0}C during sexual differentiation having masculinising effects (Baroiller and D'Cotta, 2001). Male ratios in elevated temperature-treated Nile tilapia are strongly dependent both on the population and on the parental animals (Tessema et al., 2006; Wessels & Hörstgen-Schwark, 2007). Family-specific QTL (Quantitative Trait Locus) involved in sex reversal due to temperature have been identified in LG 1, 3 and 23 in genetically all-female families, coinciding with known sex-determining regions of Nile tilapia (Lühmann et al., 2012).

Frequent departures from equal sex ratio have also been observed where the temperature was not high enough to affect sexual differentiation, and it has been postulated that these departures are caused by other loci, potentially including those in LG3 and LG23 (Mair et al., 1997; Karayücel et al., 2004; Shirak et al., 2006; Cnaani et al., 2008; Eshel et al., 2010, 2012). Interestingly crosses between putative

YY males and XX females often give less than to 100% male progeny predicted from a simple XX/XY system: some such crosses give close to 100% males, while others give lower proportions of males, but still significantly higher than the 50% expected from XY males (Beardmore et al., 2001). Many of the studies on sex determination in *O. niloticus* have been carried out on fish derived from Lake Manzala in Egypt, the subject of the present study, and it is clear that both non-LG 1 genes and temperature affect sex ratios in at least some families in this population.

In a previous study (Palaiokostas et al., 2013) we applied an NGS (Next Generation Sequencing)-based technique called RAD-seq (Restriction-site Associated DNA sequencing) in order to scan for sex-determining QTL in Nile tilapia families of balanced sex ratios, using a high-resolution genetic map based on SNPs. No other sex-determining region apart from the one in LG1 was detected. Interestingly the only misassigned individuals in both the mapping families and other samples used for validation were phenotypic males with the LG1 genotype expected of females, suggesting that they had undergone sex reversal.

Despite the numerous advantages of RAD-seq in allowing the discovery of thousands of SNPs in a cost-efficient and reproducible manner (Baird et al., 2008; McCormack et al., 2012; Davey et al., 2012), its main drawback lies in the limited number of samples that can be used in a single sequencing lane. Additional shortcomings of RAD-seq include the shearing step that has been shown to induce bias (Davey et al., 2012) and the limited control over the sequenced fragments, resulting on variable coverage across the samples. Peterson et al. (2012) elaborated

on the RAD-seq platform, introducing double digested Restriction-site Associated DNA sequencing (ddRAD-seq). In ddRAD-seq two restriction enzymes are used for digesting the genomic DNA replacing the random shearing step. ddRAD-seq results on sequencing of only a subset of genomic restriction digest fragments generated by cuts with both restriction enzymes that fall within the size-selection window, allowing for a more even coverage across the sequenced samples comparative to RAD-seq (Peterson et al., 2012). The last allows hundreds of individuals to be pooled in a single sequencing lane, providing higher statistical power for QTL mapping although with a reduced number of markers. Furthermore ddRAD requires less reads on achieving high confidence on SNP calling, due to the low chance of obtaining duplicate reads from the same restriction site. In the current study we used a modified version of the ddRAD-seq platform (allowing in the current study of sequencing three times the number of individuals that would be sequenced using RAD-seq) in order to scan for sex-determining QTL in both temperature-treated and normally reared Nile tilapia families with skewed sex ratios.

5.2 Materials and Methods

5.2.1 Sample collection and preparation

The fish used in this study came from the Tropical Aquarium Facilities of the Institute of Aquaculture at the University of Stirling. They originated from a population that was established in 1979 from Lake Manzala, Egypt (31°16'N, 32°12'E). All working procedures complied with the Animals Scientific Procedures

Act (1986). Fish were reared in recirculating water systems at 27-28°C, and fed on commercial trout diet (Trouw Aquaculture Nutrition, UK; manufacturer Skretting, UK). To set up the families used in this study, mature females were held in glass aquaria and eggs were manually stripped following ovulation. Milt was manually stripped from male fish and used to fertilise the eggs *in vitro*. Eggs were incubated in downwelling incubators until the larvae had absorbed the yolk sac. Fry from family 1 were then transferred to tanks in recirculating systems and reared for 3-4 months before being killed and sexed by microscopic examination of the gonads (Guerrero & Shelton, 1974). A sample of fin tissue was taken and fixed in 100% ethanol for DNA extraction. Family 2 was split at yolk sac absorption: one group of 80 fry was reared at 36°C for ten days (Wessels &Hörstgen-Schwark 2007) in a static 5 L tank, while a control group (80 fry) was reared at 28°C. The average survival of the two groups was 88% and 91% respectively. Subsequent rearing and sexing was as for families 1 (Table 5.1).

Offspring from family 1 showed significant deviations from equal sex ratio with 64% males ($P < 0.01$). 98 offspring were selected from family 1, in order to have a close to equal representation of males and females, for preparing a ddRAD library. A subset of the above family (46 offspring) was previously genotyped using microsatellites. Genotypic results for microsatellite UNH995 showed one of the paternal alleles appearing only in the male progeny (Khan, 2011). The dams of the above families originated from an isogenic XX line (Guyon et al., 2012; Palaiokostas et al., 2013), while sires were from the outbred red body coloured subpopulation. The available genome draft of *O. niloticus* is based on females from this isogenic

line (Broad Institute of MIT and Harvard genome assembly Orenil1.1). A second dd-RAD library was prepared from offspring of family 2 both from the control and the temperature treated group (120 offspring) (Table 5.1). The sex ratios of the control groups did not deviate from the expected 1:1 ($P > 0.3$) while the treated group showed significant deviation, giving 96% males ($P < 0.001$).

Table 5-1. Fish samples used for ddRAD Libraries.

ID	Sire strain	Dam strain	No. of females	No. of males	Total
Family1	Red†	Clonal	44	54	98
Family 2 (28°C)	Red†	Wild*	22	28	50
Family 2 (36°C)	Red†	Wild*	4	66	70

- "wild" refers to wild type coloration; † "red" refers to red body colour, which is controlled by a single gene

.

5.2.2 ddRAD library preparation and sequencing

The ddRAD library preparation protocol followed a modified version of the methodology described by Peterson et al. (2012). Each sample (0.1 µg DNA) was digested at 37°C for 40 minutes with *Sbf*I (recognising the CCTGCA|GG motif) and *SphI* (recognising GCATG|C motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U each enzyme per µg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 µg DNA per 50 µL reaction volume. The reactions (5 µL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual-specific combinations of P1 and P2 adapters, each with a unique 5-7 bp barcode, were ligated to the digested DNA at 22ºC for 60 minutes by adding 0.7/1 µL 100/25 nmol/L of P1 and P2 adapters respectively, 0.06 µL 100 mmol/L rATP (Promega), 0.95 µL 1× Reaction Buffer 2 (NEB), 0.05 µL T4 ligase (NEB, 2 M

U/mL) and reaction volumes made up to 8μL with nuclease-free water for each sample. Following heat inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in a single pool (for one sequencing lane) and purified. Size selection (300-600 bp) was performed by agarose gel separation and followed by gel purification and PCR amplification. A total of 100 μl of the amplified libraries (13-14 cycles) was purified using an equal volume of AMPure beads. After eluting into 20 μL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in two lanes of an Illumina HiSeq 2500 using 100 base paired-end reads (v3 chemistry). Raw reads were process using RTA 1.12.4.2 and Casava 1.6 (Illumina). The reads were deposited at the EBI Sequence Read Archive (SRA) study ERP004077.

5.2.3 Genotyping RAD alleles

Reads of low quality (score under 30, while the average quality score was 37), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotyped using Stacks software 1.02 (Catchen et al., 2011). The likelihood-based SNP calling algorithm (Hohenlohe et al., 2010) implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. A minimum stack depth of at least 20 and a maximum of 2 mismatches were allowed in a locus in

an individual and up to 1 mismatch between alleles. Polymorphic RAD-tags may contain more than one SNP, but the vast majority (over 99%) showed only two allelic versions; the very small proportion of RAD-tags with more than two alleles were excluded. All samples used for constructing the ddRAD libraries were also genotyped for SNPs *Oni*23063 and *Oni*28137 (NCBI dbSNP accession 748775085 and 748775081 respectively) previously described in Chapter 4.

5.2.4 Genetic map construction

SNPs were initially tested for segregation distortion using the *chisq* module of TMAP (Cartwright et al., 2007). Two genetic maps were constructed based on the offspring of families 1 and 2 using R/Onemap (Margarido et al., 2007). Recombination rates, allocation of markers into linkage groups and ordering were conducted using R/Onemap (functions: *rf.2pts, group, ug, rcd, record*). This package uses Hidden Markov Models (HMM) algorithms for outbred species while in parallel implements the methodology described in Wu et al. (2002), for calculating the most probable linkage phase. Linkage groups were formed using a minimum LOD value of 8. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function.

5.2.5 QTL mapping – Association analysis

QTL analysis for family 1 was performed using R/qtl (Broman & Sen, 2009). In family 1, with the dam originating from a clonal line and by inferring the most

probable phase of the genetic markers of the sire, the cross had the same properties as a backcross and was analysed as such. Family 1 was analysed in two steps. An initial QTL analysis was conducted using all the offspring, followed by analyzing only the male offspring (21 offspring) suspected to have undergone sex reversal, based on their female expected genotype for the SNPs in the major sex-determining region, and the female offspring (44 offspring). For family 2 the analysis was conducted using both R/qtl and GridQTL (Seaton et al., 2006). For R/qtl the cross was considered as a 'pseudo' backcross, analysing male and female informative markers separately. The half-sib regression model was used in the analysis conducted using GridQTL using treatment (temperature treated or control group) as a fixed effect. Models following single and multidimensional approach for detecting QTL were used (R/qtl functions: *scanone, addqtl*, *scantwo, fitqtl, stepwiseqtl,* GridQTL: scanning for single and two-QTL simultaneously). Permutation tests (10,000 permutations) were conducted in order to correct for the multiple testing, while in the case of GridQTL two levels of significance are reported based on chromosome- or genome-wide thresholds, with the detected QTL being referred to as suggestive or significant respectively (as in Churchill & Doerge 1994, Knott et al., 1998 and Baranski et al., 2010). Fisher's exact tests were used to test for significance between allelic combinations in different loci.

5.3 Results

5.3.1 ddRAD sequencing

In total, 573,801,882 raw reads (100 bases long) were produced (326,823,072 paired-end reads, EBI Sequence Read Archive (SRA) study ERP004077). After removing low quality sequences (quality score under 30), ambiguous barcodes and orphaned paired-end reads, 71.1% of the raw reads were retained (411,358,850 reads). In total 46,578 unique RAD-tags were retrieved, 15,577 for *Sbf*I and 31,001 for *Sph*I. The number of reads and RAD-tags for each sample are reported in the Appendix (Chapter 5:Table S6). The only misassigned offspring concerning the genotype for SNPs *Oni*23063 and *Oni*28137 were males appearing with the female expected genotype (Table 5.2).

Table 5-2. Genotypic information of samples used for ddRAD Libraries for SNPs *Oni*23063 and *Oni*28137 (NCBI dbSNP accession 748775085 and 748775081 respectively)

ID	Female Expected Genotype		Male Expected Genotype	
Family1	Females: 44	Males: 21	Females: 0	Males: 33
Family2 (28°C)	Females: 22	Males: 8	Females: 0	Males: 20
Family2 (36°C)	Females: 4	Males: 28	Females: 0	Males: 42

5.3.2 Genetic maps

In order to maximise the number of informative markers and minimise the amount of missing or erroneous data, we used SNP markers retrieved in at least 75% of the samples in each family, and carrying one or two SNPs. Genetic maps 1 and 2 were constructed from offspring of families 1 and 2 respectively. The linkage groups were named according to the Broad Institute of MIT and Harvard genome assembly Orenil1.1 (NCBI Assembly GCA_000188235.2). The first genetic map consisted of 641 SNPs (397 informative ones), grouped in 22 linkage groups that corresponded to 19 different chromosomes. The total length was 950 cM (Table 5.3). The above map did not contain information about chromosomes 5, 12 and 15 due to a lack of informative SNPs on those chromosomes. The second genetic map consisted of 1155 SNPs (642 female specific; 640 male specific; 903 informative ones), grouped in 23 linkage groups that corresponded in 22 different chromosomes (number expected from karyotype). The total length was 1382 cM (Table 5.4). The difference between the total numbers of SNPs in the two genetic maps is due to the fact that the first one contains information only from the sire since the dam originated from an isogenic line.

Table 5-3. Nile tilapia genetic map based on offspring from family 1

Linkage Group	No. of markers	No. of informative markers	Length (cM)
LG1	56	32	66.1
LG2a	14	11	30.3
LG2b	24	10	8.2
LG3	87	42	115.9
LG4	38	28	74.1
LG6	15	13	20.6
LG7	22	11	13
LG8a	23	16	33.8
LG8b	5	3	2.1

LG9	36	20	57.1
LG10	39	25	85.8
LG11	45	32	92.4
LG13	13	9	28.8
LG14	30	25	76.1
LG16	22	13	24.7
LG17	56	27	44.4
LG18	30	18	22.5
LG19	3	1	0
LG20	34	25	67.8
LG22a	23	19	43
LG22b	14	8	15.5
LG23	12	9	33.4
Total	**641**	**397**	**950.1**

Table 5-4. Nile tilapia genetic Map based on offspring from family 2

Linkage Group	No. of markers	No. of informative markers	Length (cM)
LG1	60	45	88.3
LG2	57	42	30.3
LG3	55	49	68.9
LG4	48	38	86.8
LG5	39	33	59.5
LG6	61	49	102.8
LG7	135	106	134.6
LG8	71	58	49.5
LG9	42	32	55.5
LG10	26	21	46.4
LG11	47	38	55
LG12	26	20	46.4
LG13	15	13	22.8
LG14	57	44	39.6
LG15	15	10	5.8
LG16	30	21	42.5
LG17a	22	18	27.5
LG17b	35	23	17.2
LG18	40	32	86.2
LG19	75	54	41
LG20	75	59	89.8
LG22	59	48	84.8
LG23	57	50	107.9
Total	**1155**	**903**	**1382**

5.3.3 QTL mapping

5.3.3.1 Family 1

QTL mapping for family 1 was conducted using R/qtl. The results from the single-QTL model for binary traits provided evidence for the existence of a QTL in LG 1, in the expected position of the major sex determining region (LOD=9.65), and a second QTL in LG 20 (LOD=4.87) (Figure 5.1). The genome-wide significance threshold for the single-QTL model had a value of LOD=2.77 (10,000 permutations; a=0.05). Explained variances of the above QTLs were estimated after running a multi-dimensional QTL model. QTL on LG 1 explained approximately 40.5% of the phenotypic variance (LOD=15.13; P < 2e-16), while QTL on LG 20 explained approximately 25% of the phenotypic variance (LOD=10.35; P=5.04e-12).

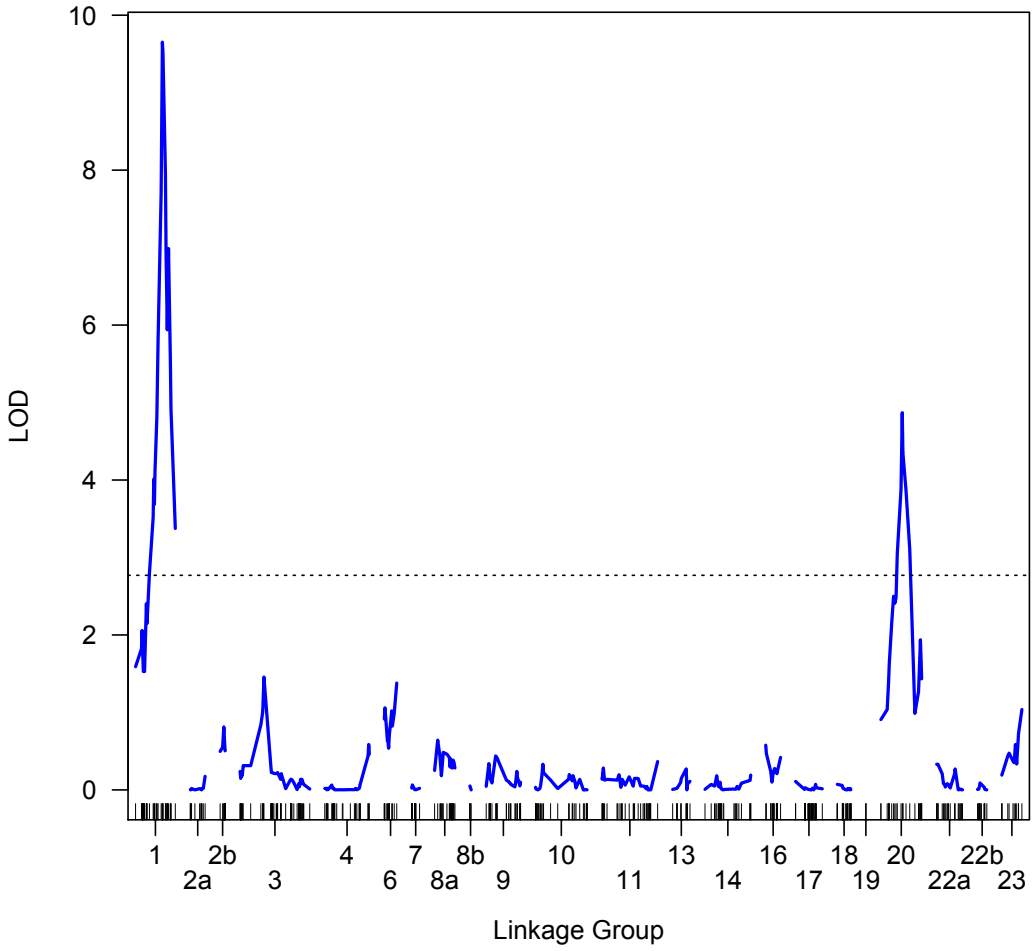

Figure 5-1. Results from QTL mapping in family 1. Plot of the LOD score (sex determining QTL search) along the linkage groups.

The estimated 95% Bayesian Density Interval for the QTL on LG 20 spanned a region of 13.5 cM (34.5-48 cM in LG 20). The QTL mapping on the reduced dataset (44 females; 21 males suspected to have undergone sex reversal) detected only the QTL on LG 20 (LOD=10.4), which explained approximately 51% of the phenotypic variance (P=5.7e-12) (Figure 5.2).

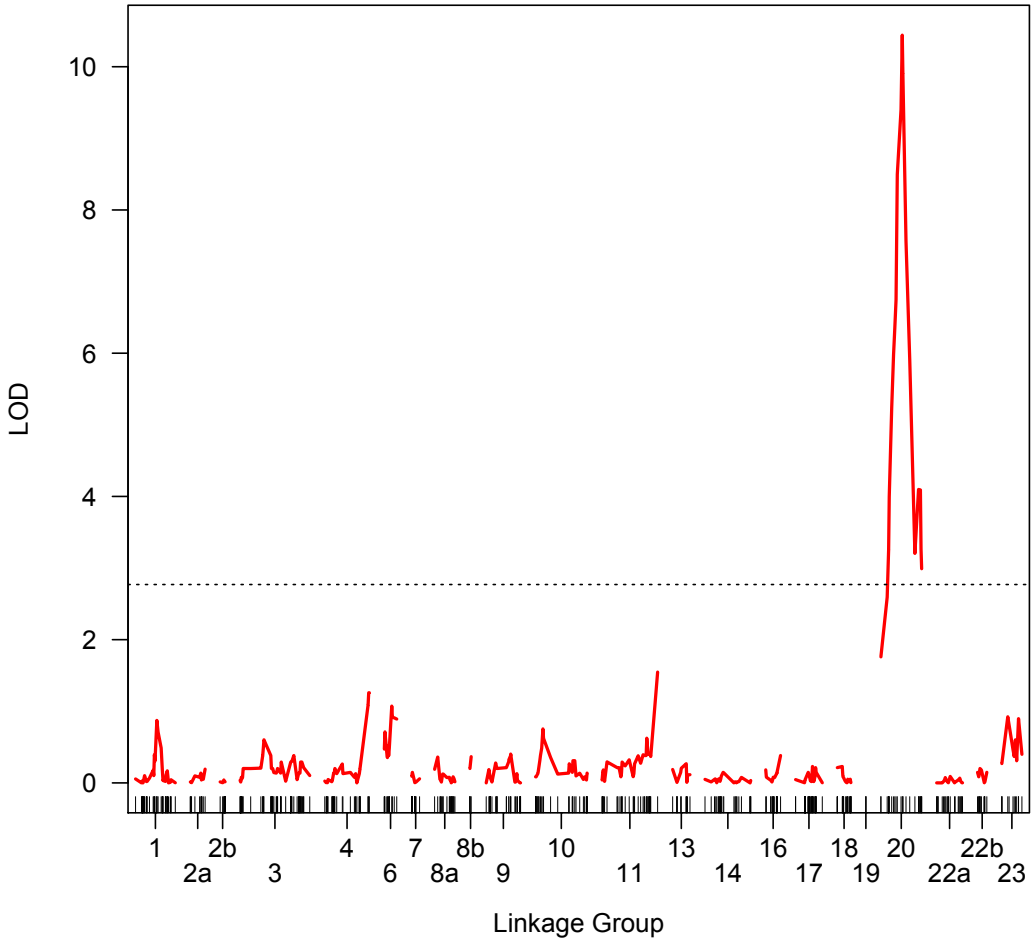

Figure 5-2. Results from QTL mapping in a reduced dataset of family 1 (44 females; 21 males suspected to have undergone sex reversal). Plot of the LOD score (sex determining QTL search) along the linkage groups.

The estimated 95% Bayesian Density Interval spanned a region of 3.5 cM (33.5-37 cM in LG 20). SNP *Oni8075* showed the highest association for the putative sex-reversed offspring. The above SNP is located in a region spanned by *Emilin-3-like* gene of *Oreochromis niloticus* (NCBI Ref: XM_003438880.2). The Fisher's exact test showed significant deviations (P=7.64e-11) in terms of allelic association in SNP *Oni8075* for offspring with the female expected genotype in the major sex

determining region, while the corresponding testing for offspring with the male expected genotype was non-significant (P > 0.95) (Table 5.5).

Table 5-5. Allelic combinations between SNPs of highest association with phenotypic sex on LG 1 (*Oni23063*) and LG 20 (*Oni8075*) for family 1 offspring

ID	Female Expected Genotype on LG 1	Male Expected Genotype on LG 1
Oni8075: AA	Females: 38 Males: 0	Females: 0 Males: 17
Oni8075: AB	Females: 8 Males: 20	Females: 0 Males: 13

5.3.3.2 Family 2

QTL mapping for family 2 was conducted using both R/qtl and GridQTL. Only male informative SNPs provided evidence for existence of sex-determining QTL. In the analysis conducted using R/qtl, the results from the single-QTL model for binary traits provided evidence for the existence of a QTL in LG 1, in the expected position of the major sex-determining region (LOD=9.66). The genome-wide significance threshold for the single-QTL model had a value of LOD=3.06 (10,000 permutations; a=0.05) (Fig. 5.3).

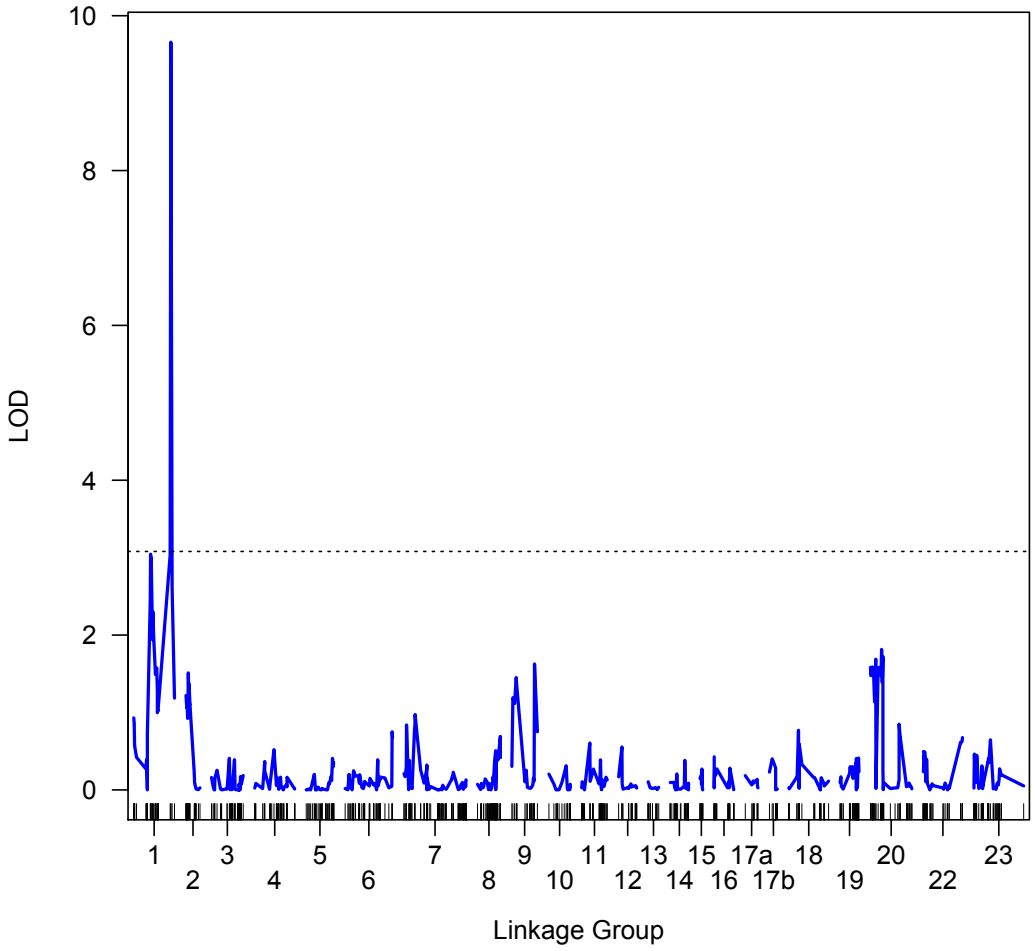

Figure 5-3. Results from QTL mapping in family 2 using R/qtl. Plot of the LOD score (sex determining QTL search) along the linkage groups.

The multidimensional QTL model provided evidence for the existence of a second QTL on LG20. In this model, the QTL on LG 1 explained approximately 28% of the phenotypic variance (LOD=9.37; P=9.6e-11), while the one on LG 20 explained approximately 6% of the phenotypic variance (LOD=2.42; P=0.0001).

In the analysis conducted using GridQTL using the half-sib regression model for single-QTL and treatment as a fixed effect, only the QTL on LG 1(F=53.62) was

significant at the genome-wide threshold, while the QTL on LG 20 (F=10.68) was significant only at the chromosome-wide level (a=0.01). The genome-wide significance threshold, estimated using 10,000 permutations had a value of F=12.28 (a=0.05). The two-QTL model indicated the possible existence of an additional QTL on LG 1 (F=17.34; comparison of the 2-QTL model with the single QTL model). The above QTL was located on the 11th cM, approximately 3.4 Mb (according to tilapia reference genome Tilapia_broad_anchored_v1) apart of the aromatase gene *cyp19a1*. The Fisher's exact tests showed significant deviations in terms of allelic combinations both in the case of the SNP with the highest association in LG 20 (*Oni8191*; P=0.014) and in the case of the SNP *Oni10909* (located on QTL detected at the beginning of LG 1; P=0.004) for offspring with the female expected genotype in the major sex determining region. The corresponding testing for offspring with the male expected genotype was non-significant (P > 0.95).

5.4 Discussion

Previous studies emphasized the complexities of Nile tilapia sex-determining system by identifying sex-determining QTL in different chromosomes (Shirak et al., 2002; Lee et al., 2003; Karayücel et al., 2004; Eshel et al., 2010). At the same time temperature also affects sex ratio, possibly interacting with genetic factors (Cnaani & Hulata, 2008) and family-specific QTL involved in temperature induced sex reversal of genotypic females have been detected (Lühmann et al., 2012). Additionally results of Ezaz et al. (2004) suggested that some genetic factors might cause sex reversal in both directions, i.e. some families showed departures from both all-male and all-female sex ratios, while others showed no departures in either direction. All the above explain to a certain degree the observed departures from equal sex ratio that a simple XY/XX sex-determining system would suggest.

The main limitation of the previous studies that tried to detect sex-determining regions in Nile tilapia was the usage of a limited number of genetic markers, mainly concentrating on linkage groups 1, 3, and 23. However, since departures from equal sex ratio are not observed in all crosses, the usage of suitable crosses is also a necessary prerequisite for mapping sex determining QTL other than LG1 in Nile tilapia, and in particular previously undetected QTL. In a previous study (Palaiokostas et al., 2013) though we managed to fine map the major sex-determining region of Nile tilapia using families with balanced sex ratios, no other sex-determining region was detected. In this study we applied ddRAD in families with skewed sex ratios, in which we expected that additional QTL would be affecting sex determination.

All the samples used in the current study were previously genotyped for two SNPs located in the expected major sex-determining region (*Oni23063, Oni28137*). Interestingly the only misassigned individuals were males appearing with the female expected genotype, following the pattern already seen in our previous study (Palaiokostas et al., 2013). This indicates sex reversal of genotypic (LG1 XX) females. The detected QTL in LG 20 in the current study provides the first evidence for a sex-determining region in this chromosome with involvement in sex reversal. The possibility that the above QTL is involved in sex reversal is strengthened also by the fact that its LOD score more than doubled in the analysis of family 1 (LOD whole dataset: 4.84; LOD reduced dataset: 10.4) when only the females and the putative sex-reversed males were used.

The QTL on LG 20 was found to be significant only at the chromosome-wide level in the first temperature treated family (Family 2). It has to be stressed though, that the temperature-treated cross was not as informative as the one of family 1. The fact that the temperature-treated group contained almost all male progeny, with only 4 females, forced us to analyze both treated and control groups simultaneously, adding the factor treatment as a fixed effect (treated; untreated). Crosses involving temperature treatment of genetic all-female progeny, as in Lühmann et al. (2012), would have been more informative. However, the fact that a signal in terms of association with sex was detected at the same genomic location in both families 1 and 2 would suggest that the same underlying mechanism is involved in both cases causing sex reversal of genotypic females.

Interestingly the 2-dimensional QTL scan in the temperature-treated family (family 2) provided evidence for another QTL at the beginning of LG 1, which could also be implicated in sex reversal. Models that take into account the existence of a major QTL (major sex determining region) or ones that test for existence of multiple QTLs simultaneously reduce the residual variation (providing higher power in the analysis for detecting additional QTLs at least of modest effect) (Broman & Sen, 2009). The location of the above QTL is in proximity (approximately 3.4 Mb) to the *cyp19a* gene. *Cyp19a* is the enzyme that catalyzes the irreversible conversion of androgens into estrogens and has been shown to be suppressed at masculinising temperatures (Penman & Piferrer, 2008). Methylation in the promoter of *cyp19a* was shown to be involved in temperature-dependent sex-determination in European sea bass (*Dicentrarchus labrax*) (Navarro-Martin et al., 2011).

5.5 Conclusions

In summary, the current study provides new insights about novel sex-determining regions in LG20 and LG1 involved in *Oreochromis niloticus* sex reversal relative to the main sex-determining region in LG1. More informative crosses, especially concerning the study of temperature effects, would shed more light concerning the magnitude of the detected QTLs. Following the hypothesis of genotypic female sex reversal it would be most interesting to study putative YY families in order to test whether the same pattern persists. The outcome of the current study stresses the complexities of the *Oreochromis niloticus* sex determining system and the potential offered by Next Generation Sequencing platforms like ddRAD-seq in unveiling those complexities.

Acknowledgments

We thank staff at The GenePool Genomics Facility, especially Urmi Trivedi and Marian Thomson, for assistance with RAD library preparation and sequencing.

Chapter 6

European sea bass

6 A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*)

Christos Palaiokostas [1], Michaël Bekaert [1], John B. Taggart [1], Karim Gharbi [2], Brendan J. McAndrew [1], David J. Penman [1,], Marc Vandeputte [3]

[1] Institute of Aquaculture, School of Natural Sciences, University of Stirling, Stirling FK9 4LA, Scotland, UK

[2] The GenePool, School of Biological Sciences, University of Edinburgh, Edinburgh EH9 3JT, Scotland, UK

[3] INRA, UMR1313 GABI Genetique animale et Biologie intgrative, 78350 Jouy-en-Josas, France

Keywords: *Dicentrarchus labrax*, Sex determination, QTL Mapping, RAD-seq, Aquaculture.

Abbreviations: RAD: restriction-site associated DNA; SNP: single nucleotide polymorphism; QTL: Quantitative trait locus; LG: Linkage group.

Status: To be submitted

Contribution: The first draft of the present manuscript was compiled and written in full by the author of this thesis, who was also fully involved in all subsequent revisions. DNA extraction, preparation of RAD libraries (under guidance of John Taggart), general statistics, construction of the linkage maps, QTL mapping, association analysis, simulation study and the genomic selection approach were carried by the candidate. The other co-authors contributed towards the experimental design, sexing, breeding of the fish, in the analysis of the sequenced reads and in the application of the machine learning algorithm.

Abstract

European sea bass (*Dicentrarchus labrax*) is one of the most important farmed species for Mediterranean aquaculture. The observed sexual dimorphism in favour of females adds additional value to deciphering its sex-determining system. Current knowledge indicates the existence of a polygenic sex-determining system interacting with temperature. RAD-seq (Restriction-site Associated DNA sequencing) was used in a test panel of 175 offspring originating from a factorial cross between two dams and four sires of one full-sib family. The current study, by making usage of restriction site associated DNA markers, provides the first SNP-based linkage maps for sea bass consisting of 2,399-3,831 SNPs grouped into 24 linkage groups, and investigates for potential quantitative trait loci (QTLs) involved in sex determination. Indications for putative sex-determining QTL, significant at the genome wide threshold, are provided in linkage groups 13 and 21. Additionally putative growth related QTL were detected in linkage groups 8, 10 and 13. A simulation run ruled out the existence of moderate to large effect sex determining QTLs explaining 20% or more of the total phenotypic variance. A suite of different models (MCMC-BLUP, BayesCPi and Bayesian Lasso) has been tested, while trying to predict phenotypic sex through the genotype, based on estimated additive single nucleotide polymorphism (SNP) effects. The MCMC-BLUP gave the best results, where only two females out of the entire dataset were misassigned, providing further indications for the existence of a polygenic sex determining system. However the pedigree based BLUP model, in which 16 mis-assignments were observed, suggested existence of over-fitting in all the above models. Usage of machine learning algorithms was

proven more efficient, where it was possible to correctly separate and assign the gender of 68% of the samples ($p < 0.0001$).

6.1 Introduction

Both genetic and environmental factors have been implicated in sex determination in fish, with some species showing influences of both factors (Devlin and Nagahama, 2002; Penman and Piferrer, 2008). Instances of both XX/XY male heterogametic and WZ/ZZ female heterogametic sex-determining systems can be found, with YY and WW genotypes being viable in most fish species tested, unlike in mammals (Volff et al., 2007; Piferrer and Guiguen, 2008).

The genetic factors could range from one or a few sex-determining quantitative trait loci (QTL) through to being truly polygenic: however, there is little evidence for polygeny. Our understanding of sex determination systems in fish has direct commercial applications as well, in a wide variety of aquaculture fish species where it relates to a range of commercially important traits like growth or age at maturity.

Dicentrarchus labrax (European sea bass) has a high commercial value with an aquaculture production of ~117,000 t in 2011. Greece and Turkey are the largest producers, with an annual production of ~43,000 t each (Federation of European Aquaculture Producers 2012). It has a production cycle of 18-24 months at the end of which it is usually marketed at around 350-400 g. Under culture conditions a high percentage of males (70-90%) is usually observed, while current evidence suggests that wild populations are possibly skewed towards females (Vandeputte et al., 2012).

Males generally grow more slowly than females, with studies showing that they could be even 40% lower in body weight at harvest time (Zanuy et al., 2001; Gorshkov et al., 2004). This dimorphism could be explained by the fact that male sea bass reach sexual maturation one year earlier than females, thus expending relatively more energy on gonad growth (Koumoundouros et al., 2002).

The karyotype of *D. labrax* consists of 24 chromosomal pairs with no recognisable heterochromosomes, although heterologous C-banding patterns on the smallest chromosome pair have been found more in male than in female *D. labrax* (Cano et al., 1996). Gynogenesis in *D. labrax* indicated that the mechanism of sex determination does not correspond to a simple monofactorial system with either male or female homogamety (Peruzzi et al., 2004). The sex ratio of the offspring from masculinised females is not female biased and would rule out both XX/XY and ZW/ZZ systems (Blázquez et al., 1999). However, sex ratio of progeny from meiogynogenetic females was skewed in favour of females (Francescon et al., 2005). *D. labrax* sex-determining system depends on both genetic factors and environmental effects with temperature being considered as the most important one under hatchery conditions (Piferrer et al., 2005; Navarro-Martin et al., 2009), with no major sex-determining gene having been identified.

Strong parental additive influences with genotype-temperature interactions can modulate the sex ratio in *D. labrax*. The proportion of females resulting from individual crossings may vary from 1 to 70% (Saillant et al., 2002; Piferrer et al., 2005). Vandeputte et al. (2007) provided the first evidence for a polygenic sex-

determining system in this species, from the analysis of between-family variation in sex ratio.

Temperature influences sex ratio in *D. labrax*, with high temperatures (>17ºC) early in development (before 60 days of age) favouring the development of males (Piferrer et al., 2005; Volckaert et al., 2008). High temperatures (~21ºC), which are typically used during the larval and early juvenile stages in aquaculture hatcheries, are thought to masculinise fish, which would have been females at lower temperatures. In the study of Navarro-Martin et al. (2009) high temperatures masculinised on average over half of the potential females. Mylonas et al. (2005) concluded that decreasing the water temperature from 21ºC to 15ºC increases the female proportion of the population, while further decreasing the rearing temperature below 15°C does not cause any further increase. However, unlike what is seen in reptiles, there is no known temperature regime, which can produce 100% males or 100% females in the sea bass. Long-term exposure to low temperatures (13°C) also has masculinising effects (Sailant et al., 2002). Additionally growth ratio at stages prior to sex differentiation has been shown to affect sex ratio (Diaz et al., 2013).

No genetic markers associated with phenotypic sex have been found until now in *D. labrax*. Restriction-site associated DNA (RAD) sequencing is a reduced-representation sequencing platform making use of high-throughput sequencing, while usage of barcodes allows multiplexing samples (Baird et al., 2008). This technique can be used for rapid discovery of thousands of single nucleotide polymorphisms (SNPs) by sequencing at high depth, offering the possibility to

construct high-density linkage maps in a cost-efficient manner. Sex-determining regions have already been identified in *Danio rerio* (zebrafish), *Oreochromis niloticus* (Nile tilapia) and *Hippoglossus hippoglossus* (Atlantic halibut) using RAD-seq (Anderson et al., 2012; Palaiokostas et al., 2013a,b). In this study we used RAD sequencing to identify SNPs in F_2 crosses of *D. labrax* that originated initially from an F_0 cross between two families of divergent sex ratio. Linkage maps were constructed, followed by QTL and association analysis with the aim of identifying sex-determining QTL and SNPs respectively.

6.2 Materials and Methods

6.2.1 Sample collection and preparation

The fish used in this study were from an F_2 population of hatchery-reared sea bass. Three F_0 males and three F_0 females, which were the offspring of wild West-Mediterranean sea bass, were mated in a factorial cross to produce 8 families (one family was lost) which were reared in common garden conditions; 43.8% of females were recorded in the F1 offspring. F_1 males and females were chosen from one family, which was a cross between the F_0 female producing the lowest proportion of females (26.6%) in its offspring and the F_0 male producing the most females (58.2%). Two F_1 females (Dam_1, Dam2) and four F_1 males (Sire_1, Sire_2, Sire_3, Sire_4) from this family were factorially mated to produce eight families (Table 6.1). In order to minimize temperature effect on sex determination, temperature was kept at 16°C from 5 to 15 dpf, 18°C from 16 to 42 dpf, and was increased to reach 25°C at 48 dpf. Temperature was stabilized between 23 and 25°C until 98 dpf, where it was

decreased to an average 21°C until the termination of the experiment. The offspring from each female (four half-sib families) were reared separately until sexing was possible at 16 months of age, when fish were killed, body length measured, weighed and sexed by visual inspection of the gonads. A fin sample from each fish was collected and stored in ethanol at room temperature. In total, 175 F_2 fish (88 males and 87 females) plus the six F_1 parents and the F_0 female were used for RAD sequencing.

6.2.2 RAD library preparation and sequencing

DNA was extracted from fin samples of the fish using the REALPure genomic DNA extraction kit (Durviz S.L.) and treated with RNase to remove residual RNA from the sample. Each sample was quantified by spectrophotometry (Nanodrop) and quality assessed by agarose gel electrophoresis, and was finally diluted to a concentration of 50 ng/μL in 5 mmol/L Tris, pH 8.5.

The RAD library preparation protocol followed essentially the methodology originally described in Baird et al. (2008) and comprehensively detailed in Etter et al. (2011), with the minor modifications described in Houston et al. (2012). The RAD-specific P1 and P2 paired-end adapters and library amplification PCR primer sequences used in this study are detailed in Baxter et al. (2011). Each sample (0.72 μg parental DNA / 0.24 μg offspring DNA) was digested at 37°C for 40 minutes with *Sbf*I (recognising the CCTGCA|GG motif) high fidelity restriction enzyme (New England Biolabs; NEB) using 6U *Sbf*I per μg genomic DNA in 1× Reaction Buffer 4 (NEB) at a final concentration of c. 1 μg DNA per 50 μL reaction

volume. The reactions (12 µL final volumes) were then heat inactivated at 65°C for 20 minutes. Individual specific P1 adapters, each with a unique 5 bp barcode (Table S1), were ligated to the *Sbf*I digested DNA at 22ºC for 60 minutes by adding 1.8 / 0.6 µL 100 nmol/L P1 adapter, 0.45 / 0.15 µL 100 mmol/L rATP (Promega), 0.75 / 0.25 µL 10× Reaction Buffer 2 (NEB), 0.36 / 0.12 µL T4 ligase (NEB, 2 M U/mL) and reaction volumes made up to 45 / 15 µL with nuclease-free water for each parental / offspring sample. Following heat inactivation at 65ºC for 20 minutes, the ligation reactions were slowly cooled to room temperature (over 1 hour) then combined in appropriate multiplex pools (Appendix: Table S1). Shearing (Covaris S2 sonication) and initial size selection (250-550 bp) by agarose gel separation (Houston et al. 2012) was followed by gel purification, end repair, dA overhang addition, P2 paired-end adapter ligation, library amplification, exactly as in the original RAD protocol (Baird et al., 2008; Etter et al., 2011). A total of 150 µL of each amplified library (14-16 PCR cycles, library dependent) was size selected (c. 350-650 bp) by gel electrophoresis (Houston et al., 2012).

Following a final gel elution step into 20 µL EB buffer (MinElute Gel Purification Kit, Qiagen), the libraries were sent to The GenePool Genomics Facility at the University of Edinburgh, UK, for quality control and high-throughput sequencing. Libraries were accurately quantified by qPCR (Kapa Library) and run in four lanes of an Illumina HiSeq 2000, using 100 base paired-end reads (v3 chemistry). Raw reads were process using RTA 1.12.4.2 and Casava 1.6 (Illumina). The reads were deposited at the EBI Sequence Read Archive (SRA) study ERP004018.

6.2.3 Genotyping RAD alleles

Reads of low quality (score under 30, while the average quality score was 37), missing the restriction site or with ambiguous barcodes were discarded. Retained reads were sorted into loci and genotypes using Stacks software 1.02 (Catchen et al., 2011). The likelihood-based SNP calling algorithm (Hohenlohe et al., 2011) implemented in Stacks evaluates each nucleotide position in every RAD-tag of all individuals, thereby differentiating true SNPs from sequencing errors. The parameters were a minimum stack depth of at least 30, a maximum of 2 mismatches allowed in a locus in an individual and up to 1 mismatch between alleles.

6.2.4 Parentage Assignment – General statistics

Parentage assignment was performed using Vitassign V8-5.1 (Vandeputte et al., 2006) using 200 SNPs. R v.3.0.1 was used for calculating descriptive statistics for each full-sib family and conducting chi-square tests for significant deviations from equal sex ratios. The above software was also used for applying a Generalised Linear Model, with phenotypic sex as a response variable and dam, sire and weight or length as explanatory variables.

6.2.5 Linkage map construction

Linkage maps based on the largest full-sib family and on informative markers for each of the two dams were constructed using R/Onemap (Margarido et al., 2007) and Tmap (Cartwright et al., 2007). SNPs were initially tested for segregation distortion using the *chisq* module of TMAP. Linkage groups were formed with a minimum

LOD value of 6. Map distances were calculated in centiMorgans (cM) using the Kosambi mapping function.

6.2.6 QTL - Association mapping

A QTL analysis, scanning for both sex-determining and growth related QTL, was performed with GridQTL (Seaton et al., 2006). Both a half-sib and a F_2 sib regression analysis model of the entire pedigree were used, allowing for the existence of both one and two QTL simultaneously. Recorded weight or length was used in the above models as a covariate. Confidence Intervals (95%) were estimated using boostraps with resampling (10,000 iterations). Two levels of significance are reported based on chromosome or genome wide thresholds, with the detected QTL being referred to as suggestive or significant respectively (as in Churchill & Doerge 1994, Knott et al., 1998 and Baranski et al., 2010).

A simulation was performed with QMSim v1.1 (Sargolzaei & Schenkel, 2009) as a measure to check the power of the performed QTL mapping for the given pedigree. A historic population with an effective population size (N_e) of 1,000 was simulated for 1,000 discrete generations. Assuming phenotypic sex as a threshold trait with an underlying distributed liability, two different lines of high and low phenotype respectively were obtained from the population above and crossed to obtain the F_2 generation of the current study. Three different data sets were simulated, with varying magnitude of the causative sex determining QTL. In the first dataset no sex determinng QTL was simulated in order to test for false positives. In the second and third simulated datesets sex determining QTL of moderate (explaining < 15% of the

phenotypic variance) and large QTL (explaining >15% of the phenotypic variance). The last two were used as approximate estimates of the statistical power of the current study. The same QTL analysis methodology was followed as with the actual dataset.

Finally an association analysis was performed using R/GenABEL (Aulchenko et al., 2007) in order to test for existing SNPs associated with sex. The last software was also used in estimating heritabilities. In all the above analysis genome-wide significance thresholds were calculated by permutation tests (10,000 permutations) in order to correct for multiple testing.

6.2.7 Prediction of phenotypic sex through estimated breeding values and machine learning algorithms

Breeding values for phenotypic sex were estimated based on additive SNP effects (4,361 SNPs) using GS3 (Legarra et al., 2008). Variance components were estimated using the MCMC (Markov Chain Monte Carlo) model of the VCE option with 1,000,000 iterations and a burn-in of the first 200,000 iterations. Prior inverted chi-square distributions with 4 df were postulated for the additive SNP variance, the polygenic variance and for the residual variance. Convergence of the resulted posterior distributions (Geweke diagnostic) and statistics of the variance components (mean, standard error) were checked using R/boa v1.1.7 (Smith, 2007). Additive SNP effects were estimated with either BayesCPi, Bayesian Lasso or MCMC-BLUP in order to check which model would give better prediction results. The tested

hypothesis was to consider positive breeding values, as corresponding to females and negative ones to males. The different families were included as a fixed effect in the models above. Additionally pedigree based breeding values were calculated using BLUP in order to check for model over-fitting issues.

The genotype of each markers for each fish examined along with their gender were entered into the WEKA package (Hall et al., 2009), which contains a variety of machine-learning algorithms, including naïve Bayes, a classifier with independent assumptions that requires a low amount of data for training to estimate parameters. Naïve Bayes is an assignment test that learns using the maximum, minimum, and mode distribution of markers in a particular gender. Naïve Bayes then classifies each individual into a particular group based on the genotype at each marker. Permutatively, one individual was removed from the training set, and subsequently the algorithm then assigned its gender. The overall score for the assignment of each individual into the right gender was obtained when all permutations were done. To measure of how different our assignment is from random, the significance of the obtained value was inferred by comparing them to the distributions of 10,000 simulated values from randomly assigned data sets (respecting the original distribution frequency).

6.3 Results

6.3.1 RAD Reads

Two crosses with 83 and 95 offspring and their parents were sampled. The samples were barcoded, pooled and sequenced in four lanes of an Illuminia HiSeq 2000 sequencer. In total, 1,156,659,542 raw reads (100 bases long) were produced (578,329,771 paired-end, EBI-SRA study ERP004018). After removing low quality sequences (quality score under 30), ambiguous barcodes and orphaned paired-end reads, 76.7% of the raw reads were retained (886,927,866 reads). The Stacks package (Catchen et al. 2011) was then used to make the assembly of the sampled loci from each individual: 56,696 unique RAD-tags were retrieved. The number of reads and RAD-tags for each sample are reported in the Appendix (Chapter 6: Table S7). In order to maximise the number of informative markers and minimise the amount of missing or erroneous data, we used RAD-tags retrieved in at least 75% of the samples in each family, and carrying one or two SNPs.

6.3.2 Parentage Assignment – General statistics

All 175 progeny (88 females and 87 males) could be assigned to a unique parental pair, by allowing a maximum of five mismatches in a test panel of 200 SNPs. Sire 403 had the lowest number of progeny while sire 141 the highest for both dams. Correlation between weight and length was 0.96. Significant sex deviations from an equal sex ratio were observed only for the parental pair of dam_2 and sire_2 (Table 6.1).

Table 6-1. Family summary -descriptive statistics – deviations from equal sex ratio.

Dam Id	*Sire Id*	*Contribution (%)*	*No Offspring*	*No. Males*	*No. Females*	*Average Weight (gr)*	*Average Length (mm)*	*X^2 (1:1 Sex ratio)*	*P-value*
Dam_1	Sire_1	7.4	13	4	9	194.2 (58.7)	256.5 (20.9)	1.23	0.27
Dam_1	Sire_2	10.8	19	14	5	179.3 (34.1)	249.5 (15.3)	3.37	0.066
Dam_1	Sire_3	9.7	17	5	12	181.9 (65.2)	248.6 (31.4)	2.12	0.15
Dam_1	Sire_4	17.7	31	17	14	178.2 (48.2)	248.7 (19.9)	0.129	0.719
Dam_2	Sire_1	9.1	16	5	11	148.1 (37)	231.5 (16.1)	1.56	0.21
Dam_2	Sire_2	12.6	22	14	8	150.2 (54.9)	233 (26.8)	1.13	0.28
Dam_2	Sire_3	10.9	19	3	16	170 (67.5)	240.2 (28.3)	7.58	0.0059**
Dam_2	Sire_4	21.7	38	25	13	160.5 (52.8)	239.6 (24.7)	3.18	0.074

a=0.05 ** a=0.01 *** a=0.001

Weight (p=0.001971), length (p=0.0007949) and sire effect (p=1.58e-05) were found as significant covariates/factor in the fitted generalized models.

6.3.3 Linkage maps

The linkage map based on the largest full-sib family (38 progeny) consisted of 3,831 SNPs that were grouped in 26 linkage groups with a total length 2,716.9 cM (Table 6.2). Four of the above linkage groups were coalesced into two (linkage groups 4a with 4b and 6a with 6b), since in the dam based linkage maps where found as overlapping. The female specific map consisted of 2,791 SNPs spanning 2,800 cM, while the male specific consisted of 2,702 SNPs spanning 2,466 cM. The two dam based genetic maps (80 and 95 progeny respectively) consisted of 2,448 and 2,399 SNPs that were grouped in 24 linkage groups in both cases, corresponding with the chromosomal number expected from karyotype, with total lengths of 2,275 and 2,797.4 cM respectively (Table 6.3, 6.4).

Table 6-2. Full-sib(Dam_2 x Sire_4) based linkage map.

LG	*No. markers*	*No. informative markers*	*Length (cM)*
1	126	109	148.1
2	194	161	143.9
3	124	114	111.6
4a	99	73	47.6
4b	69	64	63
5	50	33	29.1
6a	66	54	69.5
6b	61	58	53.7
7	196	167	126
8	247	215	123.6
9	272	242	164.3
10	180	151	126.3
11	90	77	66.2
12	138	129	68.7
13	156	142	138.8
14	71	66	53.3
15	133	117	83.3
16	114	107	143.5
17	204	182	149.2
18	154	124	70
19	196	182	133.6
20	118	111	103.5
21	262	229	130.3
22	237	196	116.1
23	99	96	115.6
24	175	139	138.1
Overall	**3,831**	**3,338**	**2,717**

Linkage map constructed using only full-sibs of the cross between Dam_2 and Sire_4.

Table 6-3. Dam_1 based linkage map.

LG	*No. markers*	*No. informative markers*	*Length (cM)*
1	137	95	108.2
2	187	155	87.3
3	102	94	122.7
4	123	113	109.6
5	137	102	113.5
6	105	89	112.8
7	120	114	116.3
8	111	104	70.1
9	115	89	104.2
10	127	99	121.2
11	79	66	91.3
12	83	71	87.8
13	110	104	110.9
14	48	48	86.4
15	116	108	82.6
16	88	62	85.2
17	77	72	100.9
18	63	54	91
19	68	61	97.1
20	56	50	84.4
21	114	93	137.7
22	66	48	28.7
23	84	55	59.9
24	132	127	65.5
Total	**2,448**	**2,073**	**2275.3**

Linkage map constructed using only Dam_1 informative SNPs across its progeny.

Table 6-4. Dam_2 based linkage map.

LG	*No. markers*	*No. informative markers*	*Length (cM)*
1	109	93	92.3
2	185	153	88.7
3	97	90	118.6
4	96	89	119.4
5	73	71	118.7
6	76	72	98.9
7	100	97	136.9
8	109	102	122.4
9	93	83	76.8
10	124	106	163.3
11	83	79	124.2
12	59	58	123.9
13	113	104	116.5
14	65	58	105
15	91	83	111.9
16	118	110	117.7
17	121	109	145.8
18	105	102	115.5
19	70	69	106.7
20	114	109	112.7
21	96	95	148.2
22	84	76	126
23	102	99	108.9
24	116	101	98.4
Total	**2,399**	**2,208**	**2,797**

Linkage map constructed using only Dam_2 informative SNPs across its progeny.

6.3.4 QTL- Association mapping

6.3.4.1 Full-sib based linkage map

The maternal half-sib regression model using weight as a covariate, detected a genome wide significant sex determining QTL at linkage group 13 (a=0.05) only in the second maternal half-sib panel (Fig 6.1A). The genome wide significant threshold was F=16.16. The confidence interval (95%) for the detected QTL spanned between the 4^{th} and the 116^{th} cM. Suggestive sex and growth related QTL were

detected in 11 different linkage groups (Table 6.5). The paternal half-sib regression models detected only suggestive sex and growth related QTL (data not shown).

Table 6-5. Mapped QTL using maternal half-sib regression analysis.

Half Sib Family	*Trait*	*LG*	*F*	*Position (cM)*
Dam_1	Sex			
		7	11.05*	64
		10	9.51*	107
		17	15.44**	123
		19	8.43*	127
	Body Weight			
		1	8.86*	102
		5	8.54*	27
		10	8.16*	82
		18	10.3*	50
	Length			
		3	7.93*	50
		10	10.74*	82
		18	9.92*	41
Dam_2				
	Sex			
		3	11.98*	74
		13	20.84***	114
		16	9.72*	103
	Body Weight			
		10	10.35*	121
		12	8.14*	54
	Length			
		10	12.03*	121
		12	9.73*	54

**** Genome wide significant ($p < 0.01$) *** Genome wide significant ($p < 0.05$) ** Chromosome wide significant ($p < 0.01$) * Chromosome wide significant ($p < 0.05$)

In the F_2 sib regression analysis, where weight was used as a covariate, a significant sex determining QTL (a=0.05) was detected at linkage group 21 (Fig 6.1B). The genome wide threshold was F=14.94. The F value of the putative QTL was 17.17 and the confidence interval (95%) spanned between the 48th and the 110th cM. Suggestive sex and growth related QTL were detected in four and two different

linkage groups respectively (Table 6.6). Using length as a covariate instead of weight did not alter the previous results.

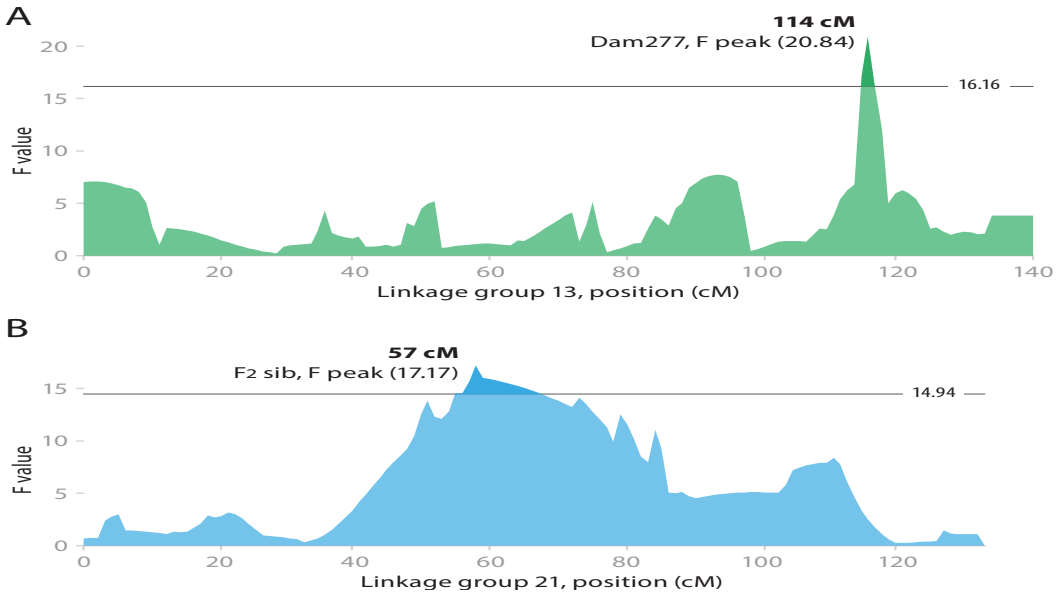

Figure 6-1. A. Sex determining QTL at linkage group 13 at the dam 277 maternal half-sibs. B. Sex determining QTL at linkage group 21 using F_2 sib regression analysis.

Table 6-6. Mapped QTL using F_2 sib regression analysis. FS (Dam_2 x Sire_4) linkage map.

Trait	*LG*	*Position*	*F*	*LOD*
Sex				
	10	53	11.47*	2.4
	13	113	9.01*	1.9
	18	70	8.06*	1.7
	21	57	17.17***	3.54
	23	30	10.71**	2.25
Weight				
	10	118	10.62*	2.23
	18	43	7*	1.49
Length				
	10	118	10.77*	2.26
	18	69	9.64*	2.03

**** Genome wide significant ($p < 0.01$)*** Genome wide significant ($p < 0.05$)** Chromosome wide significant ($p < 0.01$)* Chromosome wide significant ($p < 0.05$)

6.3.4.2 Dam_1 based linkage map

A genome wide significant growth related QTL (a=0.05) was detected at linkage 13 using the paternal half-sib regression model in the case of sire 141. The genome wide threshold was F=14.94. The F value of the putative QTL was 16.3 and the confidence interval (95%) spanned between the 34th and the 102th cM. No significant QTL was detected using the maternal half-sib regression model as opposed with the full-sib based linkage map. Suggestive sex and growth related QTL were detected generally in different linkage groups for the four paternal half-sibs (Table 6.7). With the exception of growth related QTL (based on length) that were detected between sire_1 and sire_2 (linkage group 11and 12), between sire_1 and sire_2 (linkage group 15) and finally between sire_1 and sire_4 (linkage group 13) half-sibs.

Table 6-7. Mapped QTL using paternal half-sib regression analysis. Dam_1 linkage map

Half Sib Family	*Trait*	*LG*	*F*	*Position (cM)*
Sire 1	Sex			
		8	13.05*	57
	Body Weight			
		7	11.93*	41
		9	9.35*	93
		12	9.75*	15
		13	14.3*	56
	Length			
		2	14.75*	33
		3	9.87*	8
		7	13.8*	38
		9	13.86*	93
		11	14.56*	45
		12	11.88*	16
		13	15.34*	90
		15	10.88*	42
		16	9.55*	24
		17	10.42*	36
		19	10.13*	9
		20	13.91*	40
		22	9.05*	22

		23	15.99**	42
Sire_2				
	Sex			
		19	9.62*	17
	Body Weight			
		7	11.64*	116
		9	11.4*	80
		10	12.42*	86
	Length			
		7	9.94*	116
		9	13.56*	82
		10	14.36**	85
		11	9.31*	32
		12	9.18*	16
Sire_3				
	Sex			
		2	13.71*	46
	Body Weight			
		15	11.86*	38
	Length			
		15	12.06*	38
Sire_4				
	Sex			
		11	16.98**	86
		12	12.28*	18
	Body Weight			
		13	13.22**	102
		14	8.67*	0
	Length			
		10	9.46*	95
		13	16.72***	102
		14	9.52*	0

**** Genome wide significant ($p < 0.01$)*** Genome wide significant ($p < 0.05$)** Chromosome wide significant ($p < 0.01$)* Chromosome wide significant ($p < 0.05$)

In the F_2 sib regression analysis two significant growth-related QTL (a=0.05) were detected in linkage groups 8 and 10, with F values of 17.78 and 18.13 respectively. The genome wide threshold was F=16.4. The confidence interval (95%) for the QTL at linkage group 8 spanned between the 1th and the 33th cM and for the QTL at linkage group 10 between the 53th and the 112th cM. Suggestive sex and growth related QTL were detected in two and five different linkage groups respectively (Table 6.8). The two-QTL model did not provide evidence for the existence of two QTL in any of the detected genome wide significant QTL. In the association analysis that was conducted using R/GenABEL no individual SNP was found to be

significantly associated with phenotypic sex. The estimated heritability for phenotypic sex was found to be 0.32 and for weight and length 0.55 and 0.86 respectively.

Table 6-8. Mapped QTL using F_2 sib regression analysis. Dam_2 linkage map.

Trait	*LG*	*Position*	*F*	*LOD*
Sex				
	6	61	10.39*	2.19
	11	56	16.63**	3.4
	18	78	8.66*	1.83
Weight	1	71	14.08**	2.93
	8	2	12.8**	2.67
	10	79	15.13**	3.14
	14	1	8.84*	1.87
Length				
	1	72	15.47**	3.21
	7	81	11.83*	2.48
	8	2	17.78***	3.66
	10	79	18.13***	3.73
	11	12	10.25*	2.16
	13	95	11.21*	2.35
	14	1	10.58*	2.22

**** Genome wide significant ($p < 0.01$)*** Genome wide significant ($p < 0.05$)** Chromosome wide significant ($p < 0.01$)* Chromosome wide significant ($p < 0.05$)

6.3.5 Simulation study

The anaysis of the first dataset revealed false positive suggestive QTL (significant at the chromosomal level). No genome wide significant QTL was detected when phenotypic sex was simulated as a polygenic trait only in this dataset. Simulated QTL explaining less than 15% of the total phenotypic variance could not be detected as well. QTL explaining between 15-20% of the total phenotypic variance had F values between 20.25- 51.42, surpassing the genome wide significant threshold (a=0.01).

6.3.6 Prediction of phenotypic sex through estimated breeding values

The Geweke diagnostic did not provide significant evidence of non-convergence ($p = 0.96$) for the posterior distribution of the additive SNP variance or for the rest of the variance components (data not shown). The breeding values resulting from the additive SNP effects was negative in all male progeny for all three tested models, with the exception of one offspring in the Bayesian Lasso model. Larger differences between the three models were observed in the female progeny. The MCMC-BLUP model gave the best results in terms of predicting phenotypic sex where 2 females were mis-assigned (negative estimated breeding value), while in the BayesCpi model 9 female offspring were mis-assigned. The Bayesian Lasso gave the worst prediction results where 40 female offspring were mis-assigned. Using the pedigree based estimated breeding values through BLUP 14 females and 2 males were mis-assigned.

Using the machine learning approach it was possible to obtain an assignment score significantly deviating from the assignment scores obtained from randomization (Fig. 6.2). The score for each random assignment was gathered and plotted in Figure 6.2. These results were also repeated when removing one entire family at the time. Based on Naïve Bayes assignment test, it was possible to correctly separate and assign the gender of 68% of the samples ($p < 0.0001$) in the actual dataset of the current study.

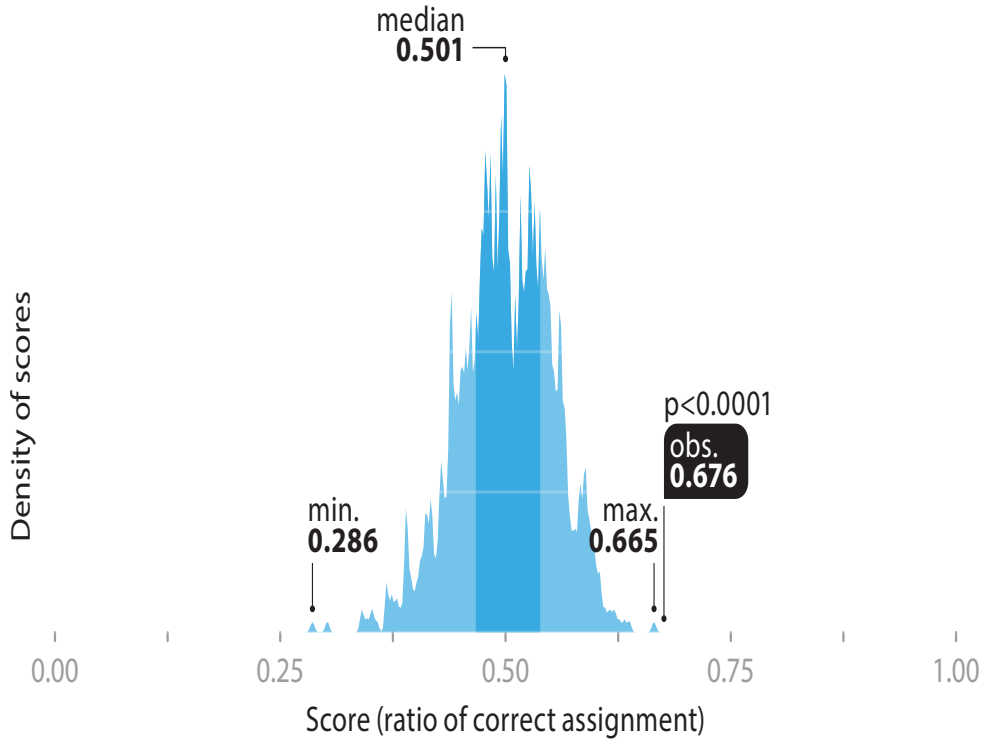

Figure 6-2. Distribution of the results of Bayesian assignment test calculated after score of correct gender assignment. The minimum, median and maximum values of the score across the randomisations are reported. The marginal lighter shades highlight the 1st and 4th quantiles. The highlighted black frame corresponds to the assignment score of the actual dataset ($p < 0.0001$).

6.4 Discussion

D. labrax is one of the most important species in Mediterranean aquaculture. The observed sexual dimorphism offers practical applications of an understanding of its sex determining system. Current evidence, based on sex-ratio variance between families, suggests the existence of a polygenic sex-determining system (Vandeputte et al., 2007). Usage of genetic markers offers the potential of directly accessing genetic variance and its distribution between eventual QTLs and polygenic background variation.

Sea bass has relatively rich genomic resources available, with its genome being soon available (R. Reinhardt, personal communication). However, available linkage maps for sea bass are based mainly only on microsatellites and AFLPs, with the most recent one consisting of 190 microsatellites, 176 AFLPs and 2 SNPs (Chistiakov et al., 2005; Chistiakov et al., 2008). RAD-seq allows for rapid discovery and genotyping of thousands of SNPs from many individuals, combined with reproducibility due to deep sequencing of the digestion fragments. RAD-seq has already been used in several fish species. Linkage maps in Zebrafish (*Danio rerio*), spotted gar (*Lepisosteus oculatus)*, Atlantic halibut (*Hippoglossus hippoglossus*) and Nile tilapia (*Oreochromis niloticus*) have been developed based on SNPs derived from RAD-seq (Anderson et al., 2012; Amores et al., 2011; Palaiokostas et al., 2013a,b).

In the present study the first linkage maps based on SNPs were constructed for sea bass. The two dam-based (half-sib) linkage maps consisted of 2,448 and 2,399 SNPs respectively, while the full-sib one (based on both parents) consisted of 3,831 SNPs. The total number of linkage groups corresponded to the number of chromosome pairs observed in sea bass. The difference in the number of SNPs between the linkage maps is due to the fact that the last one, used information from both sire and dam, while the first two used information only from the corresponding dam. These linkage maps have larger overall lengths compared to the previously published ones. The sex-averaged map produced in this study spans 2,717 cM, while that of Chistiakov et al. (2008) had a total length of 1,373 cM. This expansion is mainly due to the larger amount of markers used in the present study. The same phenomenon

was already observed between the map of Chistiakov et al. (2008) (368 markers) and the first sea bass linkage map that was based on 174 microsatellites and spanned 814 cM (Chistiakov et al., 2005).

Weight, length and sire effect were all found to be significant in the fitted generalized linear model for phenotypic sex. The fact that the dam effect was not found significant though, could be most probably due to the experimental design and the unequal sizes and sex ratios of the paternal families. Vandeputte et al. (2007) found a relatively strong genetic correlation between weight and sex ($r_A = 0.5$). The estimated heritability for sex in the present study was 0.32, while in the study of Chatziplis et al. (2007) it was estimated to be 0.12 and in the study of Vandeputte et al. (2007) to be 0.62. It has to be stressed, however, that the limited number of parents in our study prevents any attempt to generalize the heritability estimate outside the studied pedigree.

The half-sib regression model detected a genome-wide significant sex-determining QTL on linkage group 13, however this was detected only for the second half-sib group (dam 277). The F_2 sib regression analysis detected another genome-wide significant sex-determining QTL on linkage group 21. Since the latter model makes use of the entire available pedigree, its results could be considered more reliable in ruling out the possibility of a false positive. Interestingly a suggestive QTL was detected with the latter model in linkage group 13 at the same position as the QTL detected by the half-sib regression model. This represents the first indications of potential sex-determining regions in sea bass, though with low effect.

Significant growth-related QTL were detected in a number of different linkage groups (Table 5-8), adding to the existing knowledge from the previous studies of Chatziplis et al. (2007) and Massault et al. (2010). Unfortunately we were not able to test whether the putative QTL detected in our study correspond to the same genomic region with the ones from the studies above, due to the different linkage maps that were used. Interestingly we found in certain cases sex-determining and growth-related QTLs in the same linkage groups with most characteristic the case of linkage group 13, where we detected both a significant sex-determining QTL and suggestive growth related QTLs. Papadaki et al. (2005) using successive grading were able to produce a female dominant (larger fish) and a male dominant population (smaller fish), while Diaz et al. (2013) showed that a clear relation exists between growth rate at stages prior to sex differentiation (3-4 cm) and sex ratio. All the above could add to the proposed hypothesis that a minimum size is needed for the sea bass in order to differentiate as female with a co-localization in certain cases between sex-determining and growth-related QTLs.

The simulation performed ruled out the possibility of moderate to large sex-determining QTL, although leaving room for existing QTL explaining in total around 15% of the phenotypic variation, due to the statistical power limitations of the current study. The F value of the putative sex determining QTL on linkage group 21 (F = 17.17) was comparable to the ones from the simulation, with QTL explaining 15% of the phenotypic variance (F = 20.25). It has to be stressed though that QTL mapping studies, not excluding the current one, tend to overestimate the QTL effect due to the Beavis effect (Xu, 2003). The above phenomenon, that occurs due to

statistical power limitation of the conducted studies, explains also the fact that QTL are not always detected in the entire studied population, as was also observed in certain cases in the current study, but only in particular families.

The hypothesis for the existence of a polygenic sex-determining system for sea bass was supported when the efficiency of predicting phenotypic sex through the estimated breeding value of the additive SNP effects was tested. The best results were obtained using MCMC-BLUP a model that has been shown to better perform in the study of polygenic traits (Meuwissen & Goddard, 2010), while the opposite is expected in case of existing QTL with moderate to large effect where models like BayesCPi and Baysian Lasso where shown to outperform the corresponding models based on BLUP (Meuwissen et al., 2001; Kizilkaya et al., 2010; Legarra et al., 2011). Nevertheless, the observed model over-fitting constitutes dubious any attempt of deriving firm conclusions from the above. Testing the efficiency of the above models on larger datasets, so as to allow accounting for model over-fitting, would allow more firm conclusions to be drawn.

Interestingly the usage of machine learning algorithms did not seem to suffer from over-fitting issues. The results from the Naïve Bayes algorithm would support as well the hypothesis of existence of a polygenic sex determining system for European sea bass. However it has to be stressed that is of the outmost importance validating the above in additional datasets.

6.5 Conclusions

The present study clearly shows that there is not a single major sex determining gene in sea bass, as 1) we detected two QTLs and 2) a significant proportion of the sex-ratio variance remains unexplained by those QTLs. Therefore, our results, with a different approach, support the polygenic hypothesis of Vandeputte et al. (2007). Nevertheless, we still were not able to exclude the possibility of an oligogenic sex determing system due to the significant QTLs on linkage groups 13 and 21. The families used in our study originated from the West Mediterranean population, and significant differentiation exists between sea bass populations. It would seem likely that if unstable polygenic sex evolves to larger effects genes (QTLs) in sea bass, this evolution could be different in different populations. It has been shown recently that there was heterosis for sex ratio among sea bass populations (Guinand et al., in prep), so it seems rather likely that different QTLs may exist in different populations. Looking for population-specific QTLs with the same technique we used in this study would be one next logical step to unravel the complex genetic determination system of this species.

Acknowledgments

We thank staff at The GenePool Genomics Facility, especially Urmi Trivedi and Marian Thomson, for assistance with RAD library preparation and sequencing.

Chapter 7

General Discussion

7 General Discussion

The current study has attempted to elaborate and provide new insight into the complexities of the genetics of sex determination in three different farmed fish species by using techniques based on Next Generation Sequencing. An overall discussion on the main outcomes, limitations and future research directions can be given as follows:

7.1 Atlantic Halibut (*Hippoglossus hippoglossus*)

The findings presented in the current thesis confirmed the hypothesis that the Atlantic halibut has an XX/XY sex determination system. Additionally first evidence was provided concerning the location of the sex-determining region (in linkage group 13). Usually using full-sib families for QTL mapping has the advantage that linkage disequilibrium extends for larger distances compared with population data, allowing in this way to detect QTL even with less dense genotyping. The drawback of the above relies on the fact that the detected association between the genetic markers and the underlying causative effect may not persist when tested in additional families or in population data (Hayes, 2012). The fact that the association persisted in the expanded test panel (3 additional full-sib families; population data) provides evidence of strong linkage disequilibrium between the sex-associated SNPs and the causative gene(s) determining sex in this species. However, it was not possible to fine map the sex-determining region, with the calculated confidence interval spanning around 28cM. For future research it would be most interesting to genotype

large full-sib families for SNPs in this region, in order to fine map the sex-determining locus.

Interestingly, by using a variety of machine-learning algorithms 97% correct classification could be obtained. The above gives practical extensions to the outcomes of this study, since the identified SNPs could be used by the aquaculture industry assisting in the development of all-female halibut. Currently the industry relies on progeny testing for identifying neomales, a time-consuming and costly procedure that would require at least four or five years due to the timing of puberty in halibut. By employing Marker-assisted selection (MAS) it would be possible to confirm the sex-associated genotype from a non-destructive biopsy sample in hormonally-treated fish within 6-12 months of treatment, allowing neomales to be isolated and used from first maturation at three-four years post-treatment.

The current study also produced the first SNP-based genetic map for a flatfish species, with a much higher marker density than the existing halibut map. The above map could be used as a starting template or even be combined with future SNP datasets for further experiments trying to map QTL for traits of interest in halibut or other flatfish with commercial interest like sole (*Solea solea*) or turbot (*Scophthalmus maximus*).

The evidence from the current study suggests the existence of a major XX/XY sex-determining gene for this species. However, it was not possible to exclude the possibility of additional gene(s) being implicated in the process. The lack of

available genomic resources for this species and the large extension of the estimated confidence interval (22 cM) did not allow suggesting candidate major sex determining genes. However, synteny analysis showed that the sex-associated SNPs were consistently clustered in several other genomes, providing an interesting direction for further research.

7.2 Nile tilapia (*Oreochromis niloticus*)

Previous studies, mainly using microsatellite markers, showed the existence of sex-determining regions in Nile tilapia in linkage groups 1, 3 and 23 (Lee et al., 2003; Karayücel et al., 2004; Eshel et al., 2010, 2012). The main limitation of those studies were the observed inconsistencies among families concerning the association of the described genetic markers with phenotypic sex. The above underlines the main limitations of using microsatellites in studying complicated traits, such as sex determination in Nile tilapia. The difficulty of automating genotyping for a large number of microsatellites and the lower abundance compared to SNPs did not allow previous studies to carry out in-depth scans across the genome.

The current study was able to fine map the major sex-determining region in Nile tilapia in a 2cM region, which corresponded to approximately 1.2 Mb of physical distance. Most importantly the association persisted in the extended test panel (4 full-sib families and population data) explaining more than 95% of the phenotypic variance.

Additionally the current study produced the first SNP-based genetic maps for Nile tilapia. This could be of interest not only for future QTL experiments but also assisting in the ongoing genome assembly (and conversely, the draft genome was used to merge unlinked linkage groups of the corresponding genetic maps).

The availability of a draft genome assisted towards suggesting candidate major sex-determining gene(s). Interestingly the SNPs with the highest association were located in the *elongator complex protein 4-like* gene. It has to be stressed though that the current study could not exclude the possibility of an undetected major sex-determining gene. The fact that the available genome draft (Tilapia anchored v1) is based on a female means that in case the major sex-determining gene is observed only in the male, no information can be inferred by using it. Additionally it was not possible to exclude the case of no annotation concerning the major sex-determining gene.

Across the entire dataset all females were found to be homozygous for SNPs with the highest association with phenotypic sex. Males were heterozygous, with the exception of five individuals in the population and two in the family dataset. These fish possessed the homozygous genotype expected of females. The identified sex-associated SNPs could replace the need for progeny testing that is necessary until now for the production of genetically male tilapia. Interestingly no additional QTL has been detected across the test panel of full-sib families with equal sex ratio. When a family responsive to elevated temperature was genotyped for the SNPs with the highest association, the temperature treated group showed a significant number of

phenotypic males with the female compatible genotype, while the typically strong association between SNP genotype and phenotypic sex was found in the control group. The above strengthened the previous hypothesis of sex reversal of genetic females, while at the same time indicated that the genotyping of families with skewed sex ratios would be the most natural next step of research.

Genotyping two families with skewed sex ratios (the first due to unknown factors and the second due to temperature treatment) revealed an additional sex-determining QTL in linkage group 20 in both cases. This constituted the first evidence for a QTL involved in sex reversal in this linkage group. Unfortunately it was not possible to acquire a more informative cross in the case of temperature treatment than the one presented. This explains the lower statistical significance of the QTL in linkage group 20 in the temperature-treated family. However, it was not possible to exclude the possibility of different underlying mechanisms inducing sex reversal in the two different types of families that were studied.

Even though the current study successfully detected a QTL involved in the sex reversal of genotypic females, no evidence is provided as to whether the same QTL is involved in the sex reversal of genotypic males (from YY x XX crosses). Additionally it was not possible to verify whether the induced sex reversal was due to the effect of a single gene. Validating the above results in additional families where sex reversal is suspected would be most interesting. Future research work concerning genotyping fish from putative YY crosses could have high practical value as well. Detection of the sex reversal-inducing QTL in YY families would support

the hypothesis that the heavily skewed ratio towards males in those families is in fact due to sex reversal of genotypic females. However, it has to be stressed that the failure of obtaining 100% male progeny from YY males could be due to the sex reversal of genotypic males to females. All the above apart from exciting starting points for further research, have the potential of being extremely beneficial for the farmed tilapia industry.

7.3 European sea bass (*Dicentrarchus labrax*)

The first SNP-based genetic maps for sea bass were constructed in the current study. As mentioned above the constructed genetic maps could be of value in future QTL studies for traits of interest. However, it has to be stressed the small sample size especially in the case of the full-sib based genetic map. Due to this, especially the ordering of markers in each linkage group could be dubious. The mass-spawning characteristic of sea bass adds more difficulties than for the other species in keeping track of the pedigree. With the above in mind, even though a successful parentage assignment was performed in the current study using the derived SNPs, it could have been more efficient to initially use of microsatellites or ddRAD-seq in order to assign the progeny to the candidate parental pairs. This would have allowed including fewer but larger full-sib families in the studied dataset, facilitating the construction of more accurate genetic maps, especially in terms of finding the most probable order in each linkage group.

European sea bass has the most complicated sex-determining system of the studied species. Previous knowledge suggested that sea bass has a polygenic sex-determining

system (Vandeputte et al., 2007). Although the sample size used was relatively small for the study of a trait as complex as the one being studied, it was possible by using simulation to exclude the possibility of a major sex-determining gene in sea bass. Additionally the current study presents the first evidence concerning the locations of putative sex-determining regions significant at the genome wide threshold level in two different linkage groups (13, 21). Especially in the case of the QTL at linkage group 21, the F values were comparable to the ones from the simulation, with QTL explaining 15% of the phenotypic variance. It has to be stressed though that QTL mapping studies, not excluding the current one, tend to greatly overestimate the size of the QTL effect (Xu, 2003).

Interestingly in the current study in certain cases evidence of sex determining and growth-related QTLs were detected in the same linkage groups, most obviously in the case of linkage group 13. Most importantly the statistical models that were used in this study for detecting sex-determining QTL used either weight or length as a covariate. As was shown in Chapter 6 of this study the above increased the statistical power of detecting sex-determining QTL. Considering a former hypothesis, that a minimum size is needed for the sea bass in order to differentiate to female, together with a co-localization in certain cases between sex-determining and growth-related QTLs, a larger scale QTL study of genotyping large full-sib families with significant size differences and unequal sex ratios would be very interesting.

The usage of either estimated breeding values or machine-learning algorithms for predicting phenotypic sex constitutes an interesting direction for further research.

Both methodologies pointed towards the existence of a polygenic sex-determining system. However, especially in the case of estimated breeding values, the observed model over-fitting issues would make any conclusions based on the above dubious. The above phenomenon is observed in cases where the number of estimated parameters (additive SNP effects in this case) greatly surpasses the number of samples used for their estimation (number of offspring in the current case). Usage of larger datasets would allow both estimating accurate additive SNP effects on a training subset, which was not possible in the current study due to the relatively small data set, and validating the above in the rest of the test panel. This would minimize the effects of model over-fitting.

7.4 Role of Next Generation Sequencing based techniques in future developments in Aquaculture Genomics

The current study heavily relied on using molecular markers, in particular SNPs, for studying the genetics of sex determination in commercially important farmed fish. The key idea in QTL mapping is to use phenotypic data to identify regions in the genome where the genotype is statistically associated with the phenotype. However, the genotype is not observed at every possible position along the chromosomes, but only at a set of discrete landmarks called genetic markers (Broman & Sen, 2009).

The application of molecular markers in aquaculture species with the aim of improving phenotypes of interest, lags behind livestock where genomic resources are far in advance (Gjedrem & Baranski, 2009). In the case of aquaculture species genotypic information has been applied with the aim of performing Marker-assisted selection (MAS). MAS is a term used to describe the selective breeding process

where selection is practiced through the genotypes of the selection candidates (Liu & Cordes, 2004). The underlying genetic architecture of traits of interest has been treated as a "black box" in the past. The advent of molecular marker technology has allowed for a more direct selection based on the DNA level. Additional applications of molecular markers include shortening the generation interval and selection for traits that cannot be measured directly on selection candidates, such as disease resistance or meat quality traits. The difficulties of constructing pedigrees using conventional means for aquaculture species compared with livestock adds additional value to the application of molecular markers in breeding programs for parentage assignment.

Despite the initial optimism concerning the practical extensions of MAS, to date only a limited number of successful applications in aquaculture species are available. Among the most successful ones are the cases of IPN resistance in Atlantic salmon (Houston et al., 2008) and lymphocystis resistance in Japanese flounder, *Paralichthys olivaceus* (Fuji et al., 2007).

Successful application of MAS relies on the existence of strong linkage disequilibrium between the genetic markers and the underlying causative gene(s). It has been suggested that linkage disequilibrium should be at least moderate (>0.20 ; Calus et al., 2008). The above explains the limited success of MAS since most of the aquaculture species until recently at least, had limited available genomic resources. It has to be stressed that linkage disequilibrium varies between organisms, making the required number of markers vary between different species. For instance, moderate

linkage disequilibrium (> 0.20) in humans extends for around 5 Kb, in Holstein cattle for around 100 Kb and in Zhikong scallop for approximately 10 Kb (Bao, 2011).

Advances in sequencing technology have allowed many more organisms to be sequenced, including economically important species in aquaculture such as Atlantic salmon and Nile tilapia (Ragavendram & Muir, 2011). The advent of Next Generation Sequencing resulted in the rapid decline of genotypic costs making possible the application of genome-wide genomic information in animal breeding, known as genomic evaluation or selection (introduced in the paper of Meuwissen et al., 2001). The main advantages of genomic evaluation over traditional methods of estimating breeding values reside on the higher accuracy due to improved prediction of the Mendelian sampling term and the shortening of the generation interval.

As mentioned already RAD-seq and ddRAD-seq methodologies by combining control over the fragments that result from the digestion(s) with deep sequencing across individuals, allows the detection of reproducible SNPs in the magnitude of thousands (McCormack et al 2012). Though the current study, strictly speaking, could not qualify as a genome-wide study, the flexible nature of the RAD-seq and ddRAD-seq together with the diminishing costs of applying Next Generation Sequencing, would facilitate sequencing and merging datasets restricted with a wider number of restriction enzymes. This would give a more complete picture of genetic variation across the entire genome for the organism under study.

Especially in farmed species with limited genomic resources, RAD-seq and ddRAD-seq offer an excellent and economic way for conducting large-scale QTL studies that

would have been unrealistic in terms of cost for most research groups in the recent past. As already mentioned sample size is of critical value. With the above in mind, the application of ddRAD-seq using multiple sequencing lanes with different combinations of restriction enzymes each would be most interesting for future research.

According to Massault et al. (2008) it is possible to design an experiment that would have an 80% power to detect a QTL of moderate effect (explaining between 1.5 and 5% of the trait variation) by genotyping 1000 or fewer individuals. The results of all the designs explored in the above study was that having fewer larger families is more powerful for QTL detection than using a large number of families with fewer offspring. However, selecting only a few large families carries the risk that none of the parents is heterozygous for the QTL(s). Therefore the balance between family size and number of families needs careful consideration.

Applying RAD-seq using pooled DNA offers an interesting direction for future research. For the above it would be crucial to quantify initial DNA template using more accurate methodologies (e.g. using a fluorometer) than the ones used during this study. Designing appropriate algorithms that would allow estimation of the population size of the pooled samples based on the number of derived reads and assignment of individual alleles would be crucial. The above experimental design most probably would suit better studies where the phenotype of interest is of binary nature like phenotypic sex or disease resistance. Sonesson et al. (2010) presented a general DNA pooling scheme, most applied to aquaculture species where communal rearing is used. A high accuracy of selection, i.e. 0.60-0.88, was obtained with

20,000-50,000 test individuals but it was reduced when only 2,000 test individuals were used. This shows the importance of having large numbers of phenotypic records to accurately estimate marker effects. Interestingly the accuracy of selection decreased with increasing numbers of families per group.

7.5 General Summary

In summary, the main outcomes of the current study were the following:

- Construction of the first SNP-based genetic maps for Atlantic halibut, Nile tilapia and European sea bass.
- Successfully mapping the sex-determining region of Atlantic halibut and Nile tilapia.
- Provided first evidence for existence of QTL involved in sex reversal in linkage group 20 in Nile tilapia.
- Ruled out the existence of a major sex-determining gene in European sea bass.
- Provided first evidence concerning the locations of putative sex-determining QTL in European sea bass.

References

8 References

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. doi:10.1016/S0044-8486(98)00489-X.

Abucay, J.S., G.C. Mair. 1997. Hormonal sex reversal of tilapias: implications of hormone treatment application in closed water systems. *Aquaculture Research*, 28: 841-845.

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J. 1990. Basic local alignment search tool. *J Mol Biol*, 215:403–410.

Amores, A., Catchen, J.M., Ferrara, A., Fontenot, Q., Postlethwait, J.H. 2011. Genome evolution and meiotic maps by massively parallel DNA sequencing: spotted gar, an outgroup for the teleost genome duplication. *Genetics*, 188:799–808.

Anderson, J. L., Rodriguez Mari, A., Braasch, I., Amores, A., Hohenlohe, P. A., 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*(7), e40701. doi:10.1371/journal.pone.0040701.

Aulchenko, Y.S., Ripke, S., Isaacs, A., van Duijn, C.M. 2007. GenABEL: an R library for genome-wide association analysis. *Bioinformatics*, 23:1294–1296.

Babiak, J., Babiak, I., van Nes, S., Harboe, T., Haugen, T., Norberg, B. 2012. Induced sex reversal using an aromatase inhibitor, Fadrozole, in Atlantic halibut (*Hippoglossus hippoglossus L.)*. *Aquaculture*, *324-325*, 276–280.

Baird, N.A., Etter, P.D., Atwood, T.S., Currey, M.C., Shiver, A.L., Lewis, Z.A., Selker, E.U., Cresko, W.A., Johnson, E.A. 2008. Rapid SNP discovery and genetic mapping using sequenced RAD markers. *PloS ONE* 3:e3376.

Bao, Z. 2011. Comparison of Index Selection, BLUP, MAS and whole genome selection. In: Next Generation Sequencing and whole genome selection in Aquaculture. Wiley-Blackwell pg 185-218.

Baranski, M., Moen, T., Vage, D.I. 2010. Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (Salmo *salar*). *Genetics Selection Evolution* 42:17.

Baras, E., Jacobs, B., Melard, C. 2001. Effect of water temperature on survival, growth and phenotypic sex of mixed (XX-XY) progenies of Nile tilapia *Oreochromis niloticus*. *Aquaculture* 192: 187-199

Baroiller, J.F., D'Cotta, H., Bezault, E., Wessels, S., Hoerstgen-Schwark, G. 2009. Tilapia sex-determination: Where temperature and genetics meet. *Comparative Biochemistry and Physiology, Part A*.

Baroiller, J.-F., D'Cotta, H. 2001. Environment and sex-determination in farmed fish. *Comparative Biochemistry and Physiology. C: Comparative Pharmacology*, 130: 399-409.

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

Barske, L.A., Capel, B. 2008. Blurring the edges in vertebrate sex-determination. *Curr Opin Genet Dev*, 18:499–505.

Baxter SW, Davey JW, Johnston JS, Shelton AM, Heckel DG, Jiggins CD, Blaxter ML. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. *PLoS One* 6: e19315.

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

Bekaert, M. 2012. Genetic-Mapper:GNU GPL v3.

Bezault, E., Clota, F., Derivaz, M., Chevassus, B. Baroiller, J-F. 2007. Sex-determination and temperature-induced sex differentiation in three natural populations of Nile tilapia (*Oreochromis niloticus*) adapted to extreme temperature conditions. *Aquaculture* 272S1: S3-S16.

Bjornsson, B. 1995. The growth pattern and sexual maturation of Atlantic halibut (*Hippoglossus hippoglossus* L) reared in large tanks for 3 years. *Aquaculture*, 138(1-4), 281–290.

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

Blazquez, M., Felip, A., Zanuy, S., Carrillo, M., Piferrer, F. 2001. Critical period of androgen-inducible sex differentiation in a teleost fish, the European sea bass. *Journal of Fish Biology* 58:342-358.

Broman, K.W., Sen, S. 2009. A Guide to QTL Mapping with R/qtl. Springer pp:7-16.

Brown, N.P., Bromage, N.R., Penman, D.J., Shields, R.J. 1997. The karyotype of the Atlantic halibut, *Hippoglossus hippoglossus* (Linnaeus). *Aquac Research*, 28:489–491.

Bull, J. J. *Evolution of Sex Determining Mechanisms*. 1983. Menlo Park, CA: Benjamın-Cummings.

Bull, J.J., Vogt RC. 1979. Temperature-dependent sex-determination in turtles. *Science*, 206: 1186–1188.

Calus, M.P.L., Meuwissen, T.H.E., de Roos, A.P.W., Veerkamp, R.F. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics*, 170: 553-561.

Cartwright, D.A., Troggio, M., Velasco, R., Gutin, A. 2007. Genetic mapping in the presence of genotyping errors. *Genetics* 176: 2521–2527.

Catchen, J.M., Amores, A., Hohenlohe, P.A., Cresko, W.A., Postlethwait JH. 2011. Stacks: Building and genotyping loci *de novo* from short-read sequences. *G3*, 1:171–182.

Cerda, J., Douglas, S., Reith, M. 2010. Genomic resources for flatfish research and their applications. *Journal of Fish Biology*, 77: 1045-1070.

Chatain, B., Saillant, E., Peruzzi, S. 1999. Production of monosex male populations of European seabass, *Dicentrarchus labrax* L. by use of the synthetic androgen 17a-methyldehydrotestosterone. *Aquaculture* 178:225-234.

Chatziplis, D., Batargias, C., Tsigenopoulos, C.S., Magoulas, A., Kollias, S., Kotoulas, G., Volckaert, F.A.M., Haley, C.S. 2007. Mapping quantitative trait loci in European sea bass (*Dicentrarchus labrax*): The BASSMAP pilot study. *Aquaculture* 272S1:S172-S182.

Chistiakov, D.A., Tsigenopoulos, C.S., Lagnel, J., Guo, Y.-M., Hellemans, B., Haley, C.S., F.A.M. Volckaert, Kotoulas, G. 2008. A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Animal Genetics*, 39: 623-634.

Chistiakov, D.A., Hellemans, B., Haley, C.S., Law, A.S., Tsigenopoulos, C.S., Kotoulas, G., Bertotto, D., Libertini, A., Volckaert, F.A.M. 2005. A mircosatellite linkage map of the European sea bass *Dicentrarchus labrax*. *Genetics* 170, 1821-1826.

Churchill, G.A., Doerge, R.W. 1994. Empirical threshold values for quantitative trait mapping**.** *Genetics* 138:963-971.

Cnaani, A., Hulata, G. 2008a. Tilapias. In Genome Mapping and Genomics in Fishes and Aquatic Animals.Eds: Kocher, T.D., Kole, C. Springer Berlin Heidelberg Publishing pp 101-116.

Cnaani, A., Lee, B.-Y., Zilberman, N., Ozouf-Costaz, C., Hulata, G., Ron, M., D'Hont, A., Baroiller, J.-F., D'Cotta, H., Penman, D.J., Tomasino, E., Coutanceau, J.-P., Pepey, E., Shirak, A., Kocher, T.D. 2008b. Genetics of sex-determination in Tilapiine species. *Sex Dev*. 2:43-54.

Cnaani, A., Kocher, T.D. 2008c. Sex-linked markers and microsatellite locus duplication in the cichlid species *Oreochromis tanganicae*. *Biol Lett*, 4:700-703.

Cnaani, A., B.-Y. Lee, Ozouf-Costaz, C., Bonillo, C., Baroiller, J.F., D'Cotta, H., Kocher, T.D. 2007. Mapping of sox2 and sox14 in Tilapia (*Oreochromis* spp.). *Sex Dev* 1: 207-210.

Cnaani, A., Zilberman, N., Tinman, S., Hulata, G. 2004. Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. *Mol Gen Genomics* 272:162-172.

Cohen, W.,W. 1995. Fast Effective Rule Induction. In *Twelfth International Conference on Machine Learning*. Tahoe City, California, USA 115–123.

Conover, D. O. Temperature-dependent sex-determination in fishes. 2004. In: *Temperature-Dependent Sex-determination in Vertebrates* (N. Valenzuela and V. Lance, Eds.).Washington, DC: Smith-sonian Books, pp. 11–20.

Davey, J.W., Cezard, T., Fuentes-Utrilla, P., Eland, C., Gharbi, K., Blaxter, M.L. 2012. Special features of RAD Sequencing data: implications for genotyping. *Molecular Ecology* doi : 10.111/mec.12084.

Davey, J.W., Hohenlohe, P.A., Etter, P.D., Boone, J.Q., Catchen, J.M., Blaxter, M.L. 2011. Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nat Rev Genet* 17;12(7):499-510.

Deloffre, L.A.M., Martins, R.S.T., Mylonas, C.C., Canario, A.V.M. 2009. Alternative transcripts of DMRT1 in the European sea bass: Expression during gonadal differentiation. *Aquaculture* 293: 89-99.

Devlin, R.H., Nagahama, Y. 2002. Sex-determination and sex differentiation in fish: an overview of genetic, physiological, and environmental influences. *Aquaculture* 208:191-366.

Devlin, R.H., McNeil, B.K., Groves, T.D.D., Donaldson, E.M. 1991. Isolation of a Y-chromosomal DNA probe capable of determining genetic sex in Chinook salmon (*Oncorhynchus tshawytscha*). *Can J Fish Aquat Sci*, 48:1606–1612.

Diaz, N., Ribas, L., Piferrer, F. 2013. The relationship between growth and sex differentiation in the European sea bass (*Dicentrarchus labrax*). *Aquaculture* 408-409: 191-202.

Dikel S., Celik, M. 1998. Body and nutritional composition of tilapia (Tilapia spp.) from the southern seyhan river. *Turk J Vet Anim Sci* 22:517–520.

Directorate General for Health and Consumers: Directive 2003/74/EC of the European Parliament and of the Council. *Official Journal of the European Union* 2003, L 262:17–21.

El-Sayed A-F.M, Mansour, C.R., Ezzat, A.A. 2003. Effects of dietary protein level on spawning performance of Nile tilapia (*Oreochromis niloticus*) broodstock reared at different water salinities. *Aquaculture* 220:619–632.

Eshel, O. et al., 2012. Linkage and physical mapping of sex region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics*, 2(1), pp.35–42.

Eshel, O. et al., 2010. Fine-mapping of a locus on linkage group 23 for sex-determination in Nile tilapia (*Oreochromis niloticus*). *Animal genetics*, 42(2), pp.222–224.

Etter, P.D., Bassham, S., Hohenlohe, P.A., Johnson, E.A., Cresko, W.A. 2011. SNP discovery and genotyping for evolutionary genetics using RAD sequencing. *Methods Mol Biol*, 772:157–178.

Ezaz, T., Stiglec, R., Veyrunes, F., Graves, J.A. 2006. Relationships between vertebrate ZW and XY sex chromosome systems. *Current Biology*, 16:736-743.

Ezaz, M.T., Myers, J.M., Powell, S.F., McAndrew, B.J., Penman, D.J. 2004. Sex ratios in the progeny of androgenetic and gynogenetic YY male Nile tilapia, *Oreochromis niloticus* L. *Aquaculture* 232:205-214.

FAO Fisheries and Aquaculture Department (2012) FishStat database. Available: http://faostat.fao.org/site/629/default.aspx.

Fjalestad, K.T., Moen, T., Gomez-Raya, L. 2003. Prospects for genetic technology in salmon breeding programms. *Aquaculture Research* 34, 397-406.

Ferguson-Smith, M. 2006. The Evolution of sex shromosomes and sex-determination in vertebrates and the key role of DMRT1. *Sex Dev*, 1:2-11.

Flicek P, Amode MR, Barrell D, Beal K, Brent S, Carvalho-Silva D, Clapham P, Coates G, Fairley S, Fitzgerald S, Gil L, Gordon L, Hendrix M, Hourlier T, Johnson N, Kahari AK, Keefe D, Keenan S, Kinsella R, Komorowska M, Koscielny G, Kulesha E, Larsson P, Longden I, McLaren W, Muffato M, Overduin B, Pignatelli M, Pritchard B, Riat HS, et al.: Ensembl 2012. *Nucleic Acids Res* 2012, 40(Database issue):D84–90.

Ferguson, M.W., Joanen, T. 1982. Temperature of egg incubation determines sex in *Alligator mississippiensis*. *Nature*, 296:850–853.

Francescon, A., Barbaro, A., Bertotto, D., Libertini, A., Cepollaro, F., Richard, J., Belvedere, P., Colombo, L. 2005. Assessment of homozygosity and fertility in meiotic gynogens of the European sea bass (*Dicentrarchus labrax* L.). *Aquaculture* 243: 93-102.

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

Fuji, K., Hasegawa, O., Honda, K., Kumasaka, K., Sakamoto, T., Okamoto, N. 2007. Marker assisted breeding of a lymphocystis disease-resistant Japanese flounder (Paralichthys olivaceus). *Aquaculture*, 272:291-295.

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

Georges, A., Ezaz, T., Quinn, A.E., Sarre, S.D. 2010. Are reptiles predisposed to temperature-dependent sex-determination? *Sex Dev* 4: 7-15.

Gilbert H, Le Roy P, Moreno C, Robelin D, Elsen JM. 2008. QTLMAP, a software for QTL detection in outbred populations. Rotterdam, Netherlands: Annals of Human Genetics 72:694–694.

Gilbert, S. F. 2005. Mechanisms for the environmental regulation of gene expression: Ecological aspects of animal development. *J. Biosci.,* 30: 65–74.

Gjedrem, T., Baranski, M. 2009. Selective Breeding in Aquauculture: An Introduction. Springer pg 151-167.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica*, 136:245–257.

González, J.R., Armengol, L., Solé, X., Guinó, E., Mercader, J.M., Estivill, X., Moreno, V. 2007. SNPassoc: an R package to perform whole genome association studies. *Bioinformatics*, 23:644–645.

Gorshkov, S., Gorshkova, G., Meiri, I., Gordin, H. 2004a. Culture performance of different strains and crosses of the European sea bass (*Dicentrarchus labrax*) reared under controlled conditions at Eilat, Israel. *J. Appl. Ichthyol*. 20:194-203.

Gorshkov, S., Gorshkova, G., Colorni, B. 2004b. Effects of natural estradiol-17β and synthetic 17a-Ethynylestradiol on direct feminization of European sea bass *Dicentrarchus labrax*. *Journal of the World Aquaculture Society* 35:167-177.

Guerrero, R.D., Shelton, W.L. 1974. An aceto-carmine squash method for sexing juvenile fishes. *The Progressive Fish-Culturist* 36: 56–56.

Guyon, R., Rakotomanga, M., Azzouzi, N., Coutanceau, J-P., Bonillo, C., et al. 2012. A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* 13. doi:10.1186/1471-2164-13-22.

Guyon, R., Senger, F., Rakotomanga, M., Sadequi, N., Volckaert, F.A.M., Hitte, C., Galibert, F. 2010. A radiation map of the European sea bass (*Dicentrarchus labrax*) based on 1581: synteny analysis with model fish genomes. *Genomics* 96:228-238.

Hall, M., Frank, E., Holmes, G., Pfahringer, B., Reutemann, P., Witten, I.H. 2009. The WEKA data mining software. *ACM SIGKDD Explorations*, 11:10–18.

Hattori, R.S., Gould, R.J., Fujioka, T., Saito, T., Kurita, J., Strussmann, C.A., Yokota, M., Watanabe, S. 2007. Temperature-dependent sex-determination in Hd-rR Medaka *Oryzias latipes*: gender sensitivity, thermal threshold, critical period, and DMRT1 expression profile. *Sex Dev*, 1:138-146.

Harvey, S.C., Boonphakdee, C., Campos-Ramos, R., Ezaz, M.T., Griffin, D.K., Bromage, N.R., Penman, D.J. 2003a. Analysis of repetitive DNA sequences in the sex chromosomes of *Oreochromis niloticus*. *Cytogenet Genome Res*, 101:314-319.

Harvey, S.C., Kwon, J.Y., Penman, D.J. 2003b. Physical mapping of the brain and ovarian aromatase genes in the Nile tilapia, *Oreochromis niloticus*, by fluorescence in sity hybridization. *Animal Genetics*, 34:62-64.

Harvey, S.C., Campos-Ramos, R., Kennedy, D.D., Ezaz, M.T., Bromage, N.R., Griffin, D.K., Penman, D.J. 2002b. Karyotype evolution in Tilapia: mitotic and meiotic chromosome analysis of *Oreochromis karongae* and *O. niloticus* x *O. karangoe* hybrids. *Genetica*, 115:169-177.

Hayes, B., Oivind, A. 2005.Modern biotechnology and aquaculture.IN: Selection and breeding programs in Aquaculture.Ed: Gjedrem, T. Springer pp 301-319.

Hayes, B. 2012. Genomic prediction. European Institute of Statistical Genetics, Edinuburgh, UK.

Hattori, R.S., Murai, Y., Oura, M., Masuda, S., Majhi, S., K., Sakamoto, T., Fenandino, 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* 109: 2955–2959.

Hattori, R. S., R. J. Gould, T. Fujioka, T. Saito, J. Kurita, C. A. Strussmann, M. Yokota, S. Watanabe. 2007. Temperature-dependent sex-determination in *Hd-rR* medaka *Oryzias latipes*: Gender sensitivity, thermal threshold, critical period, and *DMRT1* expression profile. *Sex. Dev.,* 1: 138–146.

Hendry, C. I., Martin-Robichaud, D. J., & Benfey, T. J. 2002. Gonadal sex differentiation in Atlantic halibut. *Journal of Fish Biology*, *60*(6), 1431–1442.

Hendry, C. I., Martin-Robichaud, D. J., & Benfey, T. J. 2003. Hormonal sex reversal of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Aquaculture*, *219*(1-4), 769–781.

Hohenhohe, P.A., Bassham, S., Etter, P.D., Stiffler, N., Johnson, E.A., Cresko, W.A. 2010. Population genomics of parallel adaptation in threespine stickleback using sequenced RAD tags. *PLoS One* 6:e1000862.

Houston, R.D., Davey, J.W., Bishop, S.C., Lowe, N.R., Mota-Velasco, J.C., Hamilton, A., Guy, D.R., Tinch, A.E., Thomson, M.L., Blaxter, M.L., Gharbi, K., Bron, J.E., Taggart, J.B. 2012. Characterisation of QTL-linked and genome-wide restriction site-associated DNA (RAD) markers in farmed Atlantic salmon. *BMC Genomics*, 13:244.

Houston, R.D., Haley, C.S., Hamilton, A., Guy, D.R., Tinch, A.E., Taggart, J.B., McAndrew, B.J., Bishop, S.C. 2008. Major quantitative trait loci affect resistance to infectious pancreatic necrosis in Atlantic salmon (*Salmo salar*). *Genetics*, 178:1109–1115.

Hughes V, Benfey TJ, Martin-Robichaud DJ. 2008. Effect of rearing temperature on sex ratio in juvenile Atlantic halibut, *Hippoglossus hippoglossus*. *Environ Biol Fish*, 81:415–419.

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

Iyengar, A., Piyapattanakorn, S., Stone, D.M., Heipel, D.A, Howell, B.R., Baynes, S.M. & Maclean, N. 2000. Identification of microsatellite repeats in turbot (Scophthalmus *maximus*) and Dover Sole (*Solea solea*) using a RAPD-based technique: Characterization of microsatellite markers in Dover sole. *Mar. Biotechnol*. 2: 49-56.

Kamiya T, Kai W, Tasumi S, Oka A, Matsunaga T, et al. 2012. A trans-species missense SNP in *Amhr2* is associated with sex-determination in the tiger pufferfish, *Takifugu rubripes* (fugu). *PLoS Genetics* 8: e1002798. doi:10.1371/journal.pgen.1002798.

Karayücel İ., Ezaz T., Karayücel S., McAndrew, B.J., Penman D.J. 2004. Evidence for two unlinked "sex reversal" loci in the Nile tilapia, *Oreochromis niloticus*, and for linkage of one of these to the red body colour gene. *Aquaculture* 234: 51–63.

Kitano, J., Ross, J.A., Mori, S., Kume, M., Jones, F.C., Chan, Y.F., Absher, D.M., Grimwood, J., Schmutz, J., Myers, R.M., Kingsley, D.M., Peichel, C.L. 2009. A role for a neo-sex chromosome in stickleback speciation. *Nature* 461:1079-1083

Kizilkaya K, Fernando RL, Garrick DJ. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. *J. Anim. Sci*. 88:544–551.

Knott, S.A., Elsen, J.M., Haley, C.S. 1996. Methods for multiple marker mapping of quantitative trait loci in half-sib populations**.** *Theor Appl Genet* 93:71-8.

Koumoundouros, G., Pavlidis, M., Anezaki, L., Kokkari, C., Sterioti, A., Divanach, P., Kentouri, M. 2002. Temperature sex-determination in the European sea bass, *Dicentrarchus labrax* (L. 1758) (Teleostei, Perciformes, Moronidae): Critical sensitive ontogenetic phase. *Journal of Experimental Zoology*, 292:573-579.

Kuhl, H., Beck, A., Wozniak, G., Canario A.V.M., Volckaert, F.A.M, Reinhardt, R. 2010. The European sea bass *Dicentrarchus labrax* genome puzzle: comparative BAC-mapping and low coverage shotgun sequencing. *BMC Genomics* 11:68.

Kwon, J.Y., McAndrew, B.J., Penman, DJ. 2002. Treatment with an aromatase inhibitor suppresses high-temperature feminization of genetic male (YY) Nile tilapia. *J Fish Biol* 60: 625–636.

Lee, B-Y, Coutanceau, J-P., Ozouf-Costaz, C., D'Cotta, H., Baroiller, J-F., et al. 2011. Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Mar Biotechnol* 13: 557–562.

Lee, B-Y., Kocher, T.D. 2007. Comparative Genomics and Positional Cloning. In Aquaculture Genome Technologies. Ed: Liu, Z. Blackwell Publishing pp 323-339

Lee, B-Y, Kocher, T.D. 2007. Exclusion of *Wilms tumour* (*WT1b*) and *ovarian cytochrome P450 aromatase* (*CYP19A1*) as candidates for sex-determination genes in Nile tilapia (*Oreochromis niloticus*). *Anim Genet* 38: 85–86.

Lee, B-Y., Lee, W-J., Streelman, J.T., Carleton, K.L., Howe, A.E., Hulata, G., Slettan, A., Stern, J.E., Terai, Y., Kocher, T.D. 2005. A Second-generation linkage map of tilapia (*Oreochromis* spp.) *Genetics* 170, 237-244

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

Lee, B-Y., Penman, D.J., Kocher, T.D. 2003. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked-segregant analysis. *Anim Genet* 34, 379-383

Legarra A, Robert-Granié C, Croiseau P, Guillaume F, Fritz S. 2011. Improved Lasso for genomic selection. Genet Res (Camb) 93:77–87.

Legarra A, Robert-Granié C, Manfredi E, Elsen J-M. 2008. Performance of genomic selection in mice. Genetics 180:611–618.

Liao, X., Ma, H-Y., Xu, G-B., Shao, C-W., Tian, Y-S., Ji, X-S., Yang, J-F., Chen, S-L. 2009. Construction of a genetic linkage map and mapping of a female-specific DNA marker in half-smooth tongue sole (*Cynoglossus semilaevis*). *Mar Biotechnol*, 11:699–709.

Liew, W.C., Orban, L. 2013. Zebrafish sex: a complicated affair. *Briefings in Functional Genomics*, doi:10.1093/bfgp/elt041.

Liu, F., Sun, F., Li, J., Xia, J.H., Lin, G., Tu, R.J., Yue, G.H. 2013. A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x Oreochromis spp.) and mapping of sex-determining loci. *BMC Genomics* 14:58. doi:10.1186/1471-2164-14-58

Liu ZJ, Cordes JF. 2004. DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* 238, 1–37.

Luckenbach, J.A., Borski, R.J., Daniels, H.V., Godwin, J. 2009. Sex-determination in flatfishes: Mechanisms and environmental influences. *Semin Cell Dev Biol*, 20:256–263.

Lühmann, L.M., Knorr, C., Hörstgen-Schwark, G., Wessels, S. 2012. First evidence for family-specific QTL for temperature-dependent sex reversal in Nile tilapia (*Oreochromis niloticus*). *Sex Dev* 6: 247–256.

Ma, H., Chen, S., Yang, J., Ji, X., Chen, S., Tian, Y., Bi, J. 2010. Isolation of sex-specific AFLP markers in spotted halibut (*Verasper variegatus*). *Environ Biol Fish*, 88:9–14.

Mankiewicz, J.L., Godwin, J., Holler, B.L., Turner, P.M., Murashige, R., Shamey, R., Daniels, H.V., Borski, R.J. 2013. Masculinizing effect of background color and cortisol in a flatfish with environmental sex-determination. *Integrative and Comparative Biology*, 53:755-765

Mair, G.C., Abucay, J.S., Skibinski, D.O.F., Abella, T.A., Beardmore, J.A. 1997. Genetic manipulation of sex ratio for the large-scale production of all-male tilapia, *Oreochromis niloticus*. *Can. J. Fish. Aquat. Sci.* 54:396-404.

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

Margarido, G.R., Souza, A.P., Garcia, A.A. 2007. OneMap: software for genetic mapping in outcrossing species. *Hereditas* 144: 78–79. doi:10.1111/j.2007.0018-0661.02000.x.

Martinez, P., Bouza, C., Hermida, M., Fernandez, J., Toro, M.A., Vera, M., Pardo, B.G., Millan, A., Fernandez, C., Vilas, R., Vinas, A., Sanchez, L., Felip, A., Piferrer, F., Ferreiro, I., Cabaleiro, S. 2009. Identification of the major sex-determining region of turbot (*Scophthalmus maximus*). *Genetics*, 183:1443–1452.

Massault, C., Hellemans, B., Louro, B., Batargias, C., Van Houdt, J.K.J., Canario, A., Volckaert, F.A.M., Bovenhuis, H., Haley, C., de Koning, D.J. 2009. QTL for body weight, morphometric traits and stress response in European sea bass *Dicentrarchus labrax*. *Animal Genetics* 41: 337-345.

Massault, C., Bovenhuis, H., Haley, C., Koning, D-J. 2008. QTL mapping designs for aquaculture. *Aquaculture* 285: 23-29.

Matsuda M, Nagahama Y, Shinomiya A, Sato T, Matsuda C, et al. (2002) *DMY* is a Y-specific DM-domain gene required for male development in the medaka fish. *Nature* 417: 559–563. doi:10.1038/nature751.

McCormack JE, Hird SM, Zellmer AJ, Carstens BC, Brumfield RT. 2012. Applications of next-generation sequencing to phylogeography and phylogenetics. *Mol Phylogenet Evol* 66(2):526-538.

Meuwissen T, Goddard M. 2010. Accurate prediction of genetic values for complex traits by whole-genome resequencing. *Genetics* 185:623–631.

Meuwissen TH, Hayes BJ, Goddard ME. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829.

Miller, M.R., Brunelli, J.P., Wheeler, P.A., Liu, S., Rexroad, C.E., Palti, Y., Doe, C.Q., Thorgaard, G.H. 2012. A conserved haplotype controls parallel adaptation in geographically distant salmonid populations. *Mol Ecol*, 21:237–249.

Miller, M.R., Dunham, J.P., Amores, A., Cresko, W.A., Johnson, E.A. 2007. Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Res*, 17:240–248.

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

Miyamoto, Y., Taniguchi, H., Hamel, F., Silversides, D.W., Viger, R.S.,. 2008. A GATA4/WT1 cooperation regulates transcription of genes required for mammalian sex-determination and differentiation. *BMC Mol. Biol*. 9, 44.

Mylonas, C.C., Anezaki, L., Divanach, P., Zanuy, S., Piferrer, F., Ron, B., Peduel, A., Ben-Atia, I., Gorshkov, S., Tandler, A. 2005. Influence of rearing temperature during the larval and nursery periods on growth and sex differentiation in two

Mediterranean strains of *Dicentrarchus labrax*. *Journal of Fish Biology*, 67:652-668.

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., Hornung, U., Kondo, M., Schmid, M., Schartl, M. 2003. Common Spontaneous Sex-Reversed XX males of the Medaka *Oryzias latipes*. *Genetics* 163: 245-251.

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

Navarro-Martin, L., Vinas, J., Ribas, L., Diaz, N., Gurierrez, 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 Genetics* 7(12): e1002444. doi:10.1371/journal.pgen.1002447.

Navarro-Martin, L., Blazquez, M., Vinas, J., Joly, S., Piferrer, F. 2009a. Balancing the effects of rearing at low temperature during early development on sex ratios, growth and maturation in the European sea bass (*Dicentrarchus labrax*). Limitations and opportunities for the production of highly female-biased stocks. *Aquaculture*, 296:347-358.

Navarro-Martin, L., Blazquez, M., Piferrer, F. 2009b. Masculinization of the European sea bass (*Dicentrarchus labrax*) by treatment with an androgen or aromatase inhibitor involves different gene expression and has distinct lasting effects on maturation. *General and Comparative Endocrinology* 160:3-11.

Ocalewicz, K., Mota-Velasco, J.C., Campos-Ramos, R., Penman, D.J. 2009. FISH and DAPI staining of the synaptonemal complex of the Nile tilapia (*Oreochromis niloticus*) allow orientation of the unpaired region of bivalent 1 observed during early pachytene. *Chromosome Research*, 17:773-782.

Oosterhout, C.V., Hutchinson, W.F., Wills, D.P.M., Shipley, P. 2004. Micro-Checker:software for identifying and correcting genotyping errors in microsatellite data. *Molecular Ecology Notes* 4, 535-538.

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* 30;3(7):e2837. doi: 10.1371/journal.pone.0002837.

Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, et al. 2013a. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. *PLoS One* 8(7): e68389. doi:10.1371/journal.pone.0068389.

Palaiokostas, C., Bekaert, M., Davie, A., Cowan, M.E., Oral, M., Taggart, J.B., Gharbi, K., McAndrew, B.J., Penman, D.J., Migaud, H. 2013b. Mapping the sex-determination locus in the Atlantic halibut (*Hippoglossus hippoglossus*) using RAD sequencing. *BMC Genomics 14:556*

Papadaki, M., Piferrer, F., Zanuy, S., Maingot, E., Divanach, P., Mylonas, C. 2005. Growth, sex differentiation and gonad and plasma levels of sex steroids in male- and female-dominant populations of *Dicentrarchus labrax* obtained through repeated size grading. *Journal of Fish Biology* 66:938-956.

Parliament of the United Kingdom. 1986. Animals (Scientific Procedures) Act 1986. Available: http://www.legislation.gov.uk/ukpga/1986/14.

Penman, D.J., Piferrer, F. 2008. Fish gonadogenesis. Part I: Genetic and environmental mechanisms of sex-determination. *Reviews in Fisheries Science*, 16:16-34.

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 stickleback species. *Nature* 414, 901-915

Peruzzi, S., Chatain, B., Saillant, E., Haffray, P., Menu, B., Falguiere, J-M. 2004. Production of meiotic gynogenetic and triploid sea bass, *Dicentrarchus labrax* L. 1. Peformances, maturation and carcass quality. *Aquaculture* 230:41-64.

Peterson, B.K., Weber, J.N., Kay, E.H., Fisher, H.S., Hoekstra, H.E. 2012. Double Digest RADseq: An inexpensive method for *de novo* SNP discovery and genotyping in model and non-model species. *PloS One* 7(5): e37135. doi: 10.1371/journal.pone.0037135.

Pieau, C. 1996. Temperature variation and sex-determination in reptiles. *BioEssays,* 18: 19–26.

Piferrer, F., Guiguen, Y. 2008. Fish Gonadogenesis. Part II: Molecular Biology and genomics of sex differentiation. *Reviews in Fisheries Science*, 16:35-55.

Piferrer, F., Blazquez, M., Navarro, L., Gonzalez, A. 2005. Genetic, endocrine, and environmental components of sex-determination and differentiation in the European sea bass (*Dicentrarchus labrax* L.). *General and Comperative Endocrinology* 142:102-110.

Pompini, M., Buser, A.M., Thali, M.R., Von Siebenthal, B.A., Nussle, S., Guduff, S., Wedekind, C. 2013. Temperature-induced sex reversal is not responsible for sex-

ratio distortions in grayling *Thymallus thymallus* or brown trout *Salmo trutta*. *Journal of Fish Biology* 83, 404-411.

Ragavendram, A., Muir, W.M. 2011. Genomic Selection in Aquaculture: Methods and practical considerations. In: Next generation sequencing and whole genome selection in Aquaculture. Wiley-Blackwell pg 165-183.

Recknagel, H., Elmer, K.R., Meyer, A. 2013. A hybrid genetic linkage map of two ecologically and morphologically divergent midas cichlid fishes (*Amphilophus* spp.). Obtained by massively parallel DNA sequencing (ddRADSeq). *G3* doi: 10.1534/g3.112.003897.

Reid, D.P., Smith, C.A., Rommens, M., Blanchard, B., Martin-Robichaud, D., Reith, M. 2007. A genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics*, 177:1193–1205.

Rhen, T., Schroeder, A. 2010. Molecular mechanisms of sex-determination in reptiles. *Sex Dev* 4:16-28.

Roberts, R.,B., Ser, J.R., Kocher, T.D. 2009. Sexual conflict resolved by invasion of a novel sex determiner in lake Malawi Cichlid fishes. *Sciencexpress* 10.1126.

Ross, J.A., Urton, J.R., Boland, J., Shapiro, M.D., Peichel, C.L. 2009. Turnover of sex chromosomes in the stickleback fishes (Gasterosteidae). *PloS Genetics* 5(2): e1000391. doi:10.1371/journal.pgen.1000391

Saillant, E., Fostier, A., Haffray, P., Menu, B., Chatain, B. 2003a. Saline preferendum for the European sea bass, *Dicentrarchus labrax*, larvae and juveniles: effect of salininty on early development and sex-determination. *Journal of Experimental Marine Biology and Ecology*, 287:103-117.

Saillant, E., Fostier, A., Haffray, P., Menu, B., Laureau, S., Thimonier, J., Chatain, B. 2003b. Effects of rearing density, size grading and parental factors on sex ratios

of the sea bass (*Dicentrarchus labrax* L.) in intensive aquaculture. *Aquaculture* 221:183-206.

Saillant, E., Fostier, A., Haffray, P., Menu, B., Thimonier, J., Chatain, B. 2002. Temperature effects and Genotype-Temperature Interactions on Sex-determination in the European Sea Bass (*Dicentrarchus labrax* L.). *Journal of Experimental Zoology* 292:494-505.

Sargolzaei, M. and F. S. Schenkel. 2009. QMSim: a large-scale genome simulator for livestock. *Bioinformatics,* 25: 680-681.

Sato, T., Endo, T., Yamahira, K., Hamaguchi, S., Sakaizumi, M. 2005. Induction of female-to-male sex reversal by high temperature treatment in medaka, *Oryzias latipes*. *Zoolog Sci*, 22:985–988.

Schartl, M. 2004. Sex chromosome evolution in non-mammalian vertebrates. Current opinion in *Genetics & Development* 14: 634-641.

Seaton, G., Hernandez, J., Grunchec, J.A., White, I., Allen, J., de Koning D.J., Wei, W., Berry, D., Haley, C., Knott, S. 2006. GridQTL: A Grid Portal for QTL Mapping of Compute Intensive Datasets. Belo Horizonte, Brazil.

Sekido, R., Lovell-Badge, R. 2008 . Sex-determination and SRY: down to a wink and a nudge? *Trends Genet*, 25:19-29.

Shirak, A., Seroussi, E., Cnaani, A., Howe, A.E., Domokhovsky, R., Zilberman, N., Kocher, T.D., Hulata, G., Ron, M. 2006. Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) Linkage Group 23 within quantitative trait locus regions for sex-determination. *Genetics*, 174:1573-1581.

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

Smith, B.J. 2007. boa: An R Package for MCMC Output Convergence Assessment and Posterior Inference. *Journal of Statistical Software,* 21:11.

Socorro, S., Martins, R.S., Deloffre, L., Mylonas, C.C., Canario, A.V.M. 2007. A cDNA for European sea bass (*Dicentrarchus labrax*) 11β-hydroxulase: Gene expression during the thermosensitive period and gonadogenesis. *General and Comparative Endocrinology* 150: 164-173.

Sonesson, A.K., Meuwissen, T.H.E., Goddard, M.E. 2010. The use of communal rearing of families and DNA pooling in aquaculture genomic selection schemes. *Genetics Selection Evolution*, 42:41

Sundquist, A., Ronaghi, M., Tang, H., Pevzner, P., Batzoglou, S. 2007. Whole-genome sequencing and assembly with high-throughput, short-read technologies. *PLos One* 2(5): e484. doi:10.1371/journal.pone.0000484.

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.

Tuan P.A., Mair G.C., Little, D.C., Beardmore J.,A. 1999. Sex-determination and the feasibility of genetically male tilapia production in the Thai-Chitralada strain of *Oreochromis niloticus* (L.). *Aquaculture* 173:257–269

Tuan P.A., Little D.C., Mair G.C. 1998. Genotypic effects on comparative growth performance of all-male tilapia *Oreochromis niloticus* (L.). *Aquaculture* 159:293–302.

Valenzuela, N. 2008. Sexual Development and the Evolution of sex-determination. *Sex Dev*, 2:64-72.

Valenzuela, N., D. C. Adams, F. J. Janzen. 2003. Pattern does not equal process: Exactly when is sex environmentally determined? *Am. Nat.,* 161: 676–683.

Vandeputte, M., Quillet, E., Chatain, B. 2012. Are sex ratios in wild European sea bass (*Dicentrarchus labrax*) populations biased? *Aquat. Living Resour.* 25: 77-81.

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.

Vandeputte, M., Mauger, S., Dupont-Nivet, M. 2006. An evaluation of allowing for mismatches as a way to manage for genotyping errors in parentage assignment by exclusion. Molecular Ecology Notes 6:265-267.

Volckaert, F.A.M., Batargias, C, Canario, A., Chatziplis, D., Chistiakov, D., Haley, C., Libertini, A., Tsigenopoulos, C. European Sea Bass. 2008. In Genome mapping and genomics in fishes and aquatic animals.Eds: Kocher, T.D., Kole, C. Springer Berlin Heidelberg Publishing pp 117-133.

Volff, J.-N., Nanda, I., Schmid, M., Schartl, M. 2007. Governing sex-determination in fish: regulatory putsches and ephemeral dictators. *Sex Dev*, 1:85-99.

Wessels, S, Hörstgen-Schwark, G. 2011. Temperature dependent sex ratios in selected lines and crosses with a YY-male in Nile tilapia (*Oreochromis niloticus*). *Aquaculture* 318: 79–84.

Wessels, S., Hörstgen-Schwark, G. 2007. Selection experiments to increase the proportion of males in Nile tilapia (*Oreochromis niloticus*) by means of temperature treatment. *Aquaculture* 272: S80–S87.

Wu, R., Ma, C.X., Wu, S.S., Zeng, Z.B. 2002. Linkage mapping of sex-specific differences. *Genet Res*, 79:85–96.

Yamamoto, E. 1999. Studies on sex-manipulation and production of cloned populations in hirame, *Paralichthys olivaceus* (Temminck et Schlegel). *Aquaculture*, 173:235–246.

Yano A, Guyomard R, Nicol B, Jouanno E, Quillet E, et al. 2012a. An immune-related gene evolved into the master sex-determining gene in rainbow trout, *Oncorhynchus mykiss*. *Curr Biol* 22: 1423–1428.

Yano, A., Nicol, B., Jouanno, E., Quillet, E., Fostier, A., Guyomard, R., Guiguen, Y. 2012b. The sexually dimorphic on the Y-chromosome gene (sdY) is a conserved male-specific Y-chromosome sequence in many salmonids. *Evolutionary Applications* doi: 10.1111/eva.12032.

Xu, S. 2003. Theoretical basis of the Beavis effect. Genetics 4:2259-2268.

Zanuy, S., Carrillo, M., Felip, A., Rodriguez, L., Blazquez, M., Ramos, J., Piferrer, F. 2001. Genetic, hormonal and environmental approaches for the control of reproduction in the Europena sea bass (*Dicentrarchus labrax*). *Aquaculture*, 202:187-203.

Zerbino, D.R., Birney, E. 2008. Velvet: algorithms for de novo short read assembly using de Bruijn graphs. *Genome Res*, 18:821–829.

Appendix

9 Appendix

Chapter 3

Available online: . http://www.biomedcentral.com/1471-2164/14/566/additional

Additional file 1. **Samples origin and barcode**. Details each sample used: sample ID, family, gender, barcode used, number of extract raw reads and number of RAD-tags. Format: CSV (additional_file_1.csv)

Additional file 2. Genetic maps. Ordered markers: marker ID, linkage group and position (cM). Format: CSV (additional_file_2.csv).

Additional file 3. **Marker sequences and KASP assay primers**. List of the allele specific primers and common primer designed for the allele specific PCR genotyping assay of the 10 markers as well as their NCBI dbSNP accession numbers. Format: CSV (additional_file_3.csv).

Additional file 4. Details of the KASP assay results. Genotypes of the 94 assays. Format: CSV (additional_file_4.csv).

Additional file 5. Syntenic map of the *H. hippoglossus* sex-associated region. Each vertical block represents a segment of a different chromosome/scaffold; if the markers are on the same chromosome they are on the same block. All blocks are at the same scale. The dotted lines join the same markers from one species to the next, the solid lines are used to link markers between species further away. Format: PDF (additional_file_5.pdf).

Chapter 4

Available online:

http://www.plosone.org/article/info:doi/10.1371/journal.pone.0068389#s6

Table S1. **Samples origin and barcode**. Details each sample used: sample ID, family, gender, barcode used, number of raw reads (paired-ended or single-ended) and number of RAD-tags.

Table S2. **Genetic maps**. Ordered markers: marker ID, linkage group and position (cM).

Table S3. **KASP assay primer sequences**. List of the allele specific primers and common primer designed for the allele specific PCR genotyping assay of the five markers as well as their NCBI dbSNP accession numbers.

Table S4. **Details of the KASP assay results**. Genotypes of the 280 assays.

Data S1. **Marker sequences**. Details of SNP alleles and RAD-tag allele sequences of the five markers. (FASTA format)

Data S2. **Details of the physical location of the markers and the neighbour annotated genes**. (GFF3 format)

Chapter 5

Table S5. Summary of temperature treated crosses.

Type of cross	*No. Family*	*Treatment*	*No. males*	*No. females*	*X^2 (1:1) P-value*
Red x Red Clonal Female x Oubred					
	Family 1	Control	29	1	< 0.001
		Treated	17	0	<0.001
	Family 2	Control	6	12	1
		Treated	17	19	0.868
	Family 3	Control	15	5	0.0414
		Treated	12	12	1
	Family 4	Control	7	5	0.774
		Treated	20	12	0.215
	Family 5	Control	12	14	0.845
		Treated	11	11	1
	Family 6	Control	0	41	< 0.001
		Treated	0	31	<0.001
Outbred Female x Cloanl					
	Family 7	Control	0	68	< 0.001
		Treated	0	76	< 0.001
	Family 8	Control	0	26	< 0.001
		Treated	0	24	< 0.001
	Family 9	Control	0	37	< 0.001
		Treated	0	31	< 0.001
	Family 10	Control	0	51	< 0.001
		Treated	3	36	< 0.001
	Family 11	Control	6	59	< 0.001
		Treated	1	20	< 0.001
	Family 12	Control	3	94	< 0.001
		Treated	1	19	< 0.001
	Family 13	Control	0	61	< 0.001
		Treated	0	42	< 0.001
	Family 14	Control	0	67	< 0.001
		Treated	0	52	< 0.001
	Family 15	Control	11	59	< 0.001
		Treated	7	53	< 0.001

Wild Female x Red Male					
	Family 16	Control	47	2	< 0.001
		Treated	19	2	< 0.001
	Family 17	Control	28	20	3.12
		Treated	66	4	< 0.001
	Family 18	Control	24	21	0.76
		Treated	41	23	0.03
	Family 19	Control	14	18	0.597
		Treated	9	22	0.029
	Family 20	Control	43	14	< 0.001
		Treated	29	24	0.583
Wild Female x Clonal Male	Family 21	Control	0	31	< 0.001
		Treated	0	18	< 0.001
	Family 22	Control	0	36	< 0.001
		Treated	0	31	< 0.001
	Family 23	Control	0	103	< 0.001
		Treated	0	101	< 0.001
	Family 24	Control	9	32	< 0.001
		Treated	7	36	< 0.001
	Family 25	Control	43	9	< 0.001
		Treated	13	42	< 0.001
Red Female x Wild Male	Family 26	Control	45	39	0.586
		Treated	33	51	0.063
Wild Female x Wild Male	Family 27	Control	8	6	0.791
		Treated	62	30	0.0011

Table S6. Details of each sample used: sample ID, family, gender, barcode used, number of raw reads (paired-ended) and number of RAD-tags.

Sample_ID	*Sex*	*P1 Barcode*	*P2 Barcode*	*Reads*	*RAD-tag*
Family 1					
Dam_F	female	ACTGCAC	AGCTGTG	1779224	15097
Sire1_F1_M	male	AGGACAC	GCATA	4994956	18214
Sire2_F2_M	male	ACACGCA	GCATA	6321240	21639
F1_1_F	female	TCAGA	TAGCA	1468668	15400
F1_2_F	female	GATCG	TAGCA	500030	13253
F1_3_F	female	CATGA	TAGCA	1346402	15426
F1_4_F	female	ATCGA	TAGCA	2587462	17317
F1_5_F	female	TCGAG	TAGCA	1670720	15837
F1_6_F	female	GTCAC	TAGCA	1681116	15548
F1_7_F	female	GCATT	TAGCA	2466708	16509
F1_8_F	female	TGACC	TAGCA	1657492	15560
F1_9_F	female	TGCAACA	TAGCA	1474132	15065
F1_10_F	female	CGTATCA	TAGCA	583424	12627
F1_11_F	female	CACAGAC	TAGCA	2080120	15901
F1_12_F	female	ACTGCAC	TAGCA	1876210	15749
F1_13_F	female	TCTCTCA	TAGCA	1628106	15838
F1_14_F	female	GTACACA	TAGCA	2247574	17010
F1_15_F	female	CTCTTCA	TAGCA	1306006	14984
F1_16_F	female	CTAGGAC	TAGCA	1861012	16221
F1_17_F	female	ACGTA	TAGCA	2520820	16712
F1_18_F	female	AGAGT	TAGCA	1734864	16019
F1_19_F	female	ATGCT	TAGCA	1897854	15706
F1_20_F	female	GACTA	TAGCA	2076336	16220
F1_21_F	female	CAGTCAC	TAGCA	1547886	14963
F1_22_F	female	CAGTCCA	TAGCA	2329752	15606
F1_23_F	female	ACACGCA	TAGCA	1361090	15118
F1_24_M	male	AGGACAC	TAGCA	1668794	15501
F1_25_M	male	TCAGA	AGCTGTG	805796	14111
F1_26_M	male	GATCG	AGCTGTG	911384	14341
F1_27_M	male	CATGA	AGCTGTG	1179042	14714
F1_28_M	male	ATCGA	AGCTGTG	1198914	14639
F1_29_M	male	TCGAG	AGCTGTG	1151442	14538
F1_30_M	male	GTCAC	AGCTGTG	1013588	14370
F1_32_M	male	GCATT	AGCTGTG	1650080	15711

F1_33_M	male	TGACC	AGCTGTG	2252234	16526
F1_36_M	male	TGCAACA	AGCTGTG	1773348	15830
F1_37_M	male	CGTATCA	AGCTGTG	1393052	15038
F1_42_M	male	ATGCT	AGCTGTG	1638858	15762
F1_44_M	male	GACTA	AGCTGTG	2156620	16990
F1_45_F	female	CAGTCAC	AGCTGTG	3354720	19466
F1_46_M	male	CAGTCCA	AGCTGTG	588	2
F1_47_M	male	ACACGCA	AGCTGTG	1793100	15972
F1_48_M	male	AGGACAC	AGCTGTG	1630714	15267
F1_49_F	female	TCAGA	AGTCA	1839946	16033
F1_50_M	male	GATCG	AGTCA	1747448	16170
F1_51_M	male	CATGA	AGTCA	2279116	16576
F1_52_F	female	ATCGA	AGTCA	2071060	16404
F1_53_M	male	TCGAG	AGTCA	2392400	17035
F1_54_M	male	GTCAC	AGTCA	2115616	16595
F1_55_F	female	GCATT	AGTCA	2722580	17343
F1_56_F	female	TGACC	AGTCA	2473494	16351
F1_57_M	male	TGCAACA	AGTCA	3398566	19086
F1_58_M	male	CGTATCA	AGTCA	1971136	15876
F1_59_M	male	CACAGAC	AGTCA	1786506	16013
F1_61_M	male	ACTGCAC	AGTCA	2152654	16344
F1_62_F	female	TCTCTCA	AGTCA	1385200	15190
F1_63_M	male	GTACACA	AGTCA	1828134	16034
F1_64_F	female	CTCTTCA	AGTCA	1663546	16215
F1_65_M	male	CTAGGAC	AGTCA	1696102	15891
F1_66_F	female	ACGTA	AGTCA	2406968	16612
F1_67_F	female	AGAGT	AGTCA	1849158	16218
F1_68_F	female	ATGCT	AGTCA	2228010	16694
F1_70_F	female	GACTA	AGTCA	2487664	17009
F1_71_M	male	CAGTCAC	AGTCA	2018656	16446
F1_72_F	female	CAGTCCA	AGTCA	1689510	15684
F1_73_M	male	ACACGCA	AGTCA	1989642	16444
F1_74_M	male	AGGACAC	AGTCA	2642772	17025
F1_75_M	male	TCAGA	TACGTGT	1418862	15426
F1_76_M	male	GATCG	TACGTGT	1783494	15986
F1_77_M	male	CATGA	TACGTGT	1814630	16162
F1_79_M	male	ATCGA	TACGTGT	3324250	17761
F1_80_M	male	TCGAG	TACGTGT	2005818	15292
F1_81_M	male	GTCAC	TACGTGT	3364782	19156
F1_82_M	male	GCATT	TACGTGT	3124834	17345
F1_83_M	male	TGACC	TACGTGT	3261256	16607

F1_84_F	female	TGCAACA	TACGTGT	3318772	17689
F1_85_F	female	CGTATCA	TACGTGT	2925780	16544
F1_86_M	male	CACAGAC	TACGTGT	2360020	16151
F1_87_M	male	ACTGCAC	TACGTGT	2706026	17055
F1_88_M	male	TCTCTCA	TACGTGT	1083994	14449
F1_89_F	female	GTACACA	TACGTGT	2144630	16533
F1_90_F	female	CTCTTCA	TACGTGT	2071736	15830
F1_91_M	male	CTAGGAC	TACGTGT	2411828	16060
F1_92_M	male	ACGTA	TACGTGT	1866536	15202
F1_93_F	female	AGAGT	TACGTGT	1731626	15262
F1_94_M	male	ATGCT	TACGTGT	2503264	16423
F1_95_M	male	GACTA	TACGTGT	2467786	15936
F1_96_M	male	CAGTCAC	TACGTGT	2758416	17419
F1_97_F	female	CAGTCCA	TACGTGT	2841354	16819
F1_98_F	female	ACACGCA	TACGTGT	3020492	16496
F1_99_F	female	AGGACAC	TACGTGT	3193978	16951
F1_100_M	male	TCAGA	GCATA	2322916	16510
F1_101_F	female	GATCG	GCATA	2491588	16695
F1_102_M	male	CATGA	GCATA	3066802	17850
F1_31_M	male	CACAGAC	GCATA	1746632	15367
F1_34_M	male	ACTGCAC	GCATA	1324920	14919
F1_35_M	male	TCTCTCA	GCATA	1747790	15689
F1_38_M	male	GTACACA	GCATA	1515996	15095
F1_43_F	female	ATGCT	GCATA	1482986	14807
F1_39_M	male	GACTA	GCATA	1665632	15491
F1_40_M	male	CAGTCAC	GCATA	2167588	15867
F1_78_M	male	CAGTCCA	GCATA	1861528	15191
Family 2					
F00D_M	male	TCAGA	TAGCA	1036532	14412
C4a_F	female	GATCG	TAGCA	4692162	22375
T31C_1_M	male	ACGTA	TAGCA	2235356	16353
T31C_2_M	male	AGAGT	TAGCA	2023960	17556
T31C_3_M	male	ATGCT	TAGCA	2320406	18062
T31C_4_M	male	GACTA	TAGCA	2419478	18057
T31C_5_M	male	CAGTCAC	TAGCA	1215474	14750
T31C_6_M	male	CAGTCCA	TAGCA	2221456	17734
T31C_7_M	male	ACACGCA	TAGCA	2603084	18092
T31C_8_M	male	AGGACAC	TAGCA	3253232	18833
T31C_9_M	male	TCAGA	AGCTGTG	1232304	14390
T31C_10_M	male	GATCG	AGCTGTG	1467580	15479
T31C_11_M	male	CATGA	AGCTGTG	1076140	14111

T31C_12_M	male	ATCGA	AGCTGTG	2010046	16918
T31C_13_F	female	TCGAG	AGCTGTG	1999596	16524
T31C_14_F	female	GTCAC	AGCTGTG	1475772	14723
T31C_15_F	female	GCATT	AGCTGTG	1640892	15317
T31C_16_F	female	TGACC	AGCTGTG	1986370	16338
T31C_17_F	female	TGCAACA	AGCTGTG	984680	13455
T31C_18_F	female	CGTATCA	AGCTGTG	2130854	18133
T31C_19_F	female	CACAGAC	AGCTGTG	1194634	15427
T31C_20_F	female	ACTGCAC	AGCTGTG	1662262	15851
T31C_21_F	female	TCTCTCA	AGCTGTG	1185460	14185
T31C_22_F	female	GTACACA	AGCTGTG	1397592	14798
T31C_23_F	female	CTCTTCA	AGCTGTG	1516418	15214
T31C_24_M	male	CTAGGAC	AGCTGTG	1886564	16148
T31C_25_M	male	ACGTA	AGCTGTG	1574018	15401
T31C_26_M	male	AGAGT	AGCTGTG	1195830	14893
T31C_27_M	male	ATGCT	AGCTGTG	1605234	15354
T31C_28_M	male	GACTA	AGCTGTG	1336586	14884
T31C_29_M	male	CAGTCAC	AGCTGTG	1307132	14806
T31C_30_M	male	CAGTCCA	AGCTGTG	1054444	14251
T31C_31_M	male	ACACGCA	AGCTGTG	1221318	14281
T31C_32_M	male	AGGACAC	AGCTGTG	1531678	15161
T31C_33_M	male	TCAGA	AGTCA	1999888	16748
T31C_34_M	male	GATCG	AGTCA	1151758	13952
T31C_35_M	male	CATGA	AGTCA	1284736	14244
T31C_36_M	male	ATCGA	AGTCA	2845278	18150
T31C_37_F	female	TCGAG	AGTCA	1421814	14648
T31C_38_F	female	GTCAC	AGTCA	1552196	14942
T31C_39_F	female	GCATT	AGTCA	2039662	16251
T31C_40_F	female	TGACC	AGTCA	2422926	17619
T31C_41_F	female	TGCAACA	AGTCA	2092692	17149
T31C_50_M	male	CGTATCA	AGTCA	2015410	16217
T31C_51_M	male	CACAGAC	AGTCA	1919532	16701
T31C_52_M	male	ACTGCAC	AGTCA	1998946	16711
T31C_53_F	female	TCTCTCA	AGTCA	1739114	16470
T31C_54_F	female	GTACACA	AGTCA	2028392	16742
T31C_55_F	female	CTCTTCA	AGTCA	2021000	17068
T31C_56_F	female	CTAGGAC	AGTCA	1903102	16849
T31T_1_M	male	ACGTA	AGTCA	1909020	15638
T31T_2_M	male	AGAGT	AGTCA	1357794	15303
T31T_3_M	male	ATGCT	AGTCA	1356694	14680
T31T_4_M	male	GACTA	AGTCA	1496726	15344

T31T_5_F	female	CAGTCAC	AGTCA	1235010	14433
T31T_6_M	male	CAGTCCA	AGTCA	961516	13926
T31T_7_M	male	ACACGCA	AGTCA	2077690	16795
T31T_8_M	male	AGGACAC	AGTCA	1974578	15988
T31T_9_M	male	TCAGA	TACGTGT	1034118	13698
T31T_10_M	male	GATCG	TACGTGT	1407770	14411
T31T_11_M	male	CATGA	TACGTGT	1146310	13982
T31T_12_M	male	ATCGA	TACGTGT	1391038	14339
T31T_13_M	male	TCGAG	TACGTGT	1954728	15718
T31T_14_M	male	GTCAC	TACGTGT	1187072	14214
T31T_15_M	male	GCATT	TACGTGT	1526688	14675
T31T_16_M	male	TGACC	TACGTGT	1603212	14709
T31T_17_M	male	TGCAACA	TACGTGT	1466374	14796
T31T_18_M	male	CGTATCA	TACGTGT	1605588	15105
T31T_19_M	male	CACAGAC	TACGTGT	1156590	13794
T31T_20_M	male	ACTGCAC	TACGTGT	1927924	15814
T31T_21_M	male	TCTCTCA	TACGTGT	1619652	15118
T31T_22_M	male	GTACACA	TACGTGT	1464444	14484
T31T_23_M	male	CTCTTCA	TACGTGT	1746510	15417
T31T_24_M	male	CTAGGAC	TACGTGT	1427546	14639
T31T_25_M	male	ACGTA	TACGTGT	1380624	14300
T31T_26_M	male	AGAGT	TACGTGT	1058400	14105
T31T_27_M	male	ATGCT	TACGTGT	1384552	14768
T31T_28_M	male	GACTA	TACGTGT	1018120	13675
T31T_29_F	female	CAGTCAC	TACGTGT	817286	13551
T31T_30_M	male	CAGTCCA	TACGTGT	1115154	13903
T31T_31_M	male	ACACGCA	TACGTGT	1018058	13779
T31T_32_M	male	AGGACAC	TACGTGT	1841072	15304
T31T_33_M	male	TCAGA	GCATA	2027854	16158
T31T_34_M	male	GATCG	GCATA	1545280	14663
T31T_35_M	male	CATGA	GCATA	1484582	14494
T31T_36_M	male	ATCGA	GCATA	1205302	14056
T31T_37_M	male	TCGAG	GCATA	1826666	15253
T31T_38_M	male	GTCAC	GCATA	1480394	14646
T31T_39_F	female	GCATT	GCATA	1858876	15205
T31T_40_M	male	TGACC	GCATA	2506282	16223
T31T_41_M	male	TGCAACA	GCATA	1509724	14376
T31T_42_M	male	CGTATCA	GCATA	2449682	16818
T31T_43_M	male	CACAGAC	GCATA	1621666	15263
T31T_44_M	male	ACTGCAC	GCATA	1371720	14373
T31T_45_M	male	TCTCTCA	GCATA	1289392	14437

T31T_46_M	male	GTACACA	GCATA	1450056	14727
T31T_47_M	male	CTCTTCA	GCATA	1316856	14476
T31T_48_M	male	CTAGGAC	GCATA	1616462	15280
T31T_49_M	male	ACGTA	GCATA	1819404	15116
T31T_50_M	male	AGAGT	GCATA	1548682	15401
T31T_51_F	female	ATGCT	GCATA	1568806	14774
T31T_52_M	male	GACTA	GCATA	1146456	14197
T31T_53_M	male	CAGTCAC	GCATA	1074078	13955
T31T_54_M	male	CAGTCCA	GCATA	1270758	14906
T31T_55_M	male	ACACGCA	GCATA	1228998	14058
T31T_56_M	male	AGGACAC	GCATA	2275842	16743
T31T_57_M	male	TCAGA	GAGATGT	1377728	14318
T31T_58_M	male	GATCG	GAGATGT	742556	12802
T31T_59_M	male	CATGA	GAGATGT	1053028	14005
T31T_60_M	male	ATCGA	GAGATGT	1535878	14779
T31T_61_M	male	TCGAG	GAGATGT	1708690	15231
T31T_62_M	male	GTCAC	GAGATGT	946864	13646
T31T_63_M	male	GCATT	GAGATGT	1100124	13683
T31T_64_M	male	TGACC	GAGATGT	1548284	14525
T31T_65_M	male	TGCAACA	GAGATGT	1220796	14008
T31T_66_M	male	CGTATCA	GAGATGT	1465224	15109
T31T_67_M	male	CACAGAC	GAGATGT	863070	13350
T31T_68_M	male	ACTGCAC	GAGATGT	1439908	14988
T31T_69_M	male	TCTCTCA	GAGATGT	1450970	15150
T31T_70_M	male	GTACACA	GAGATGT	1151988	13816
T31C_58_F	female	CTCTTCA	GAGATGT	1416016	15523
T31C_59_F	female	CTAGGAC	GAGATGT	1406850	15203

Chapter 6

Table S7. Details of each sample used: sample ID, family, gender, barcode used, number of raw reads (paired-ended) and number of RAD-tags.

Sample_ID	*Cross*	*Gender*	*Barcode*	*Paired-end_Reads*	*RAD-tag*
Parent_C_G	-	female	ACTGC	6676176	49119
Parent_258_D1	1	female	ACCAT	3336639	48066
Parent_277_D2	2	female	TGGTT	2194319	47568
Parent_105_S1	1 & 2	male	TAATG	2295367	47904
Parent_407_S2	1 & 2	male	TCAGA	991001	48205
Parent_403_S3	1 & 2	male	AACCC (1) & GTACA (3)	424345	39156
Parent_141_S4	1 & 2	male	TTAAT	1638806	47586
Off_1	1	female	CGATA	8516134	49884
Off_2	1	female	CTAGG	8507501	50113
Off_3	1	male	CTGAA	2973790	47349
Off_4	1	male	CGCGC	1779547	46236
Off_5	1	male	GCTAA	2459173	48224
Off_25	1	female	GGCCT	2690587	47209
Off_26	1	male	GTCAC	2127949	46608
Off_28	1	male	TAGCA	3521416	47694
Off_29	1	female	TCCTC	987677	45255
Off_31	1	male	TGTGG	2686845	47805
Off_34	1	female	GCCGG	2099813	47134
Off_36	1	female	GTGTG	5122712	48495
Off_37	1	male	CGTAT	3013847	48701
Off_39	1	female	CTCTT	1294896	46028
Off_40	1	male	CCTTG	833549	44888
Off_67	1	female	GACTA	1769319	47261
Off_68	1	male	GCGCC	8732781	54454
Off_70	1	male	GTTGT	2301536	47457
Off_62	1	female	TACGT	3384675	48404
Off_27	1	female	TCGAG	4559544	49781
Off_30	1	female	TGACC	44793	2979
Off_48	1	female	ATCGA	1581294	46674
Off_35	1	male	TTTTA	1516355	47102
Off_43	1	female	CCCCA	1353491	46888
Off_49	1	male	CGGCG	8218919	52061
Off_54	1	male	GATCG	1674373	46961

Off_55	1	male	GCATT	7334921	53533
Off_58	1	male	GGTTC	2246407	47439
Off_59	1	female	TATAC	2220512	47907
Off_60	1	male	TTCCG	2767221	48221
Off_76	1	male	AGCTG	1660253	47133
Off_80	1	female	AGTCA	1969221	47116
Off_69	1	male	CTTCC	1308006	46592
Off_50	1	female	CATGA	803208	44886
Off_38	1	female	GTACA	3242510	49075
Off_77	1	female	GAGAT	3645923	50063
Off_78	1	female	ATTAG	445738	39832
Off_79	1	male	AGAGT	15414	438
Off_46	1	female	CACAG	866590	45363
Off_65	1	male	TCTCT	1225932	45846
Off_89	1	female	TGCAA	2286066	47063
Off_90	1	male	GAAGC	9276041	52209
Off_128	1	female	GGGGA	8075011	49697
Off_121	1	male	CGATA	2462764	48750
Off_133	1	male	CTAGG	4393193	59817
Off_134	1	male	CTGAA	5048793	54213
Off_136	1	male	CGCGC	742667	44858
Off_53	1	female	TAATG	3759567	51122
Off_75	1	female	TCAGA	3470483	49631
Off_143	1	male	GTCAC	948491	45809
Off_145	1	male	TAGCA	5194559	54277
Off_255	1	female	GGTTC	1127072	46224
Off_258	1	female	TATAC	1251125	47248
Off_164	1	female	GCCGG	1159131	46338
Off_167	1	male	CGTAT	1449858	47172
Off_172	1	male	GTGTG	6188150	55653
Off_175	1	female	CTCTT	1632151	47339
Off_177	1	male	CCTTG	1358445	47094
Off_180	1	female	GACTA	1250418	46977
Off_182	1	female	GCGCC	6448180	51183
Off_183	1	male	GTTGT	2227691	49005
Off_184	1	male	TACGT	2293182	48841
Off_195	1	male	ATCGA	3031561	49259
Off_192	1	male	TGACC	13508	859
Off_187	1	female	TCGAG	4045039	50656
Off_200	1	male	TTTTA	1356499	47369
Off_203	1	female	CCCCA	1180864	46321

Off_238	1	female	CGGCG	5140333	51185
Off_239	1	male	GATCG	1113795	46491
Off_245	1	male	GCATT	5752775	53140
Off_150	1	male	TCCTC	1121423	46997
Off_295	1	female	AGCTG	1685535	47326
Off_296	1	female	AGTCA	1525796	46659
Off_157	1	male	TGTGG	975093	45936
Off_261	1	male	TTCCG	1493233	47164
Off_52	1	female	CTTCC	2432059	49031
Off_140	1	male	GCTAA	950302	46076
Off_142	1	male	GGCCT	1462282	47785
Off_92	1	female	AACCC	4318489	52010
Off_98	1	female	ACCAT	6864861	54585
Off_101	1	female	ACTGC	5621547	54038
Off_113	1	female	TGGTT	4480910	52713
Off_114	1	female	TTAAT	2489842	53757
Off2_2	2	female	CGATA	5034496	49311
Off2_3	2	female	CTAGG	4663165	51575
Off2_4	2	female	CTGAA	4194656	50190
Off2_5	2	male	CGCGC	1261866	46034
Off2_7	2	male	GCTAA	1665723	46980
Off2_13	2	male	GGCCT	1771871	47009
Off2_14	2	male	GTCAC	1563017	46683
Off2_15	2	female	TAGCA	6493068	50646
Off2_17	2	male	TCCTC	1370996	46322
Off2_18	2	male	TGTGG	2321953	47926
Off2_19	2	female	GCCGG	1872374	47310
Off2_24	2	female	CGTAT	1092624	46463
Off2_21	2	male	GTGTG	2343022	47430
Off2_25	2	female	CTCTT	455161	41348
Off2_26	2	female	CCTTG	553448	43058
Off2_29	2	female	GACTA	745494	44258
Off2_32	2	male	GCGCC	2732850	48558
Off2_34	2	male	GTTGT	705935	44398
Off2_37	2	female	TACGT	971413	45424
Off2_44	2	male	TCGAG	2089141	47574
Off2_51	2	male	GAGAT	2454676	47714
Off2_61	2	female	ATCGA	775632	44685
Off2_67	2	male	CCCCA	491856	42333
Off2_79	2	female	AGTCA	509473	42248
Off2_66	2	male	TTTTA	1149109	46088

Off2_68	2	male	CGGCG	3207780	49240
Off2_69	2	female	GATCG	857425	45484
Off2_70	2	male	GCATT	3995727	49804
Off2_72	2	male	GGTTC	1145605	46089
Off2_73	2	female	TATAC	1319204	46697
Off2_75	2	female	TTCCG	1494067	46300
Off2_78	2	male	AGCTG	1082804	45510
Off2_87	2	male	CTTCC	977676	46232
Off2_88	2	female	CATGA	1250133	46940
Off2_95	2	female	ATTAG	707042	44427
Off2_101	2	female	CACAG	463463	41594
Off2_93	2	male	TCTCT	395230	40407
Off2_98	2	male	TGCAA	830881	45034
Off2_104	2	female	GAAGC	2280891	46957
Off2_107	2	male	GGGGA	1620818	47039
Off2_111	2	female	TAATG	1920300	46567
Off2_112	2	female	TCAGA	1473443	46632
Off2_114	2	female	AACCC	2288096	46952
Off2_115	2	female	ACCAT	1773975	46489
Off2_125	2	female	ACTGC	1825782	46817
Off2_136	2	female	GGAAG	2056427	46747
Off2_141	2	male	TTGGC	1457903	47305
Off2_145	2	male	CGATA	4355731	51789
Off2_146	2	male	CTAGG	4360526	53132
Off2_151	2	female	CTGAA	5067916	53768
Off2_166	2	male	CGCGC	735212	44637
Off2_190	2	male	GCTAA	1007958	45807
Off2_197	2	male	GGCCT	802287	45030
Off2_208	2	female	GTCAC	894382	45573
Off2_269	2	female	TAGCA	2233862	48364
Off2_273	2	female	TCCTC	520097	41029
Off2_222	2	female	TGTGG	1345885	46753
Off2_223	2	female	GCCGG	1027638	45959
Off2_227	2	male	CGTAT	1088596	46332
Off2_229	2	male	GTGTG	2885162	54203
Off2_232	2	male	CTCTT	1506188	47226
Off2_235	2	female	CCTTG	1390872	46792
Off2_238	2	male	GACTA	1575162	47109
Off2_239	2	female	GCGCC	3165606	50735
Off2_243	2	female	GTTGT	1689003	47401
Off2_251	2	male	TACGT	2152747	47659

Off2_257	2	male	TCGAG	4346975	56601
Off2_258	2	male	GAGAT	3665324	52966
Off2_260	2	male	ATCGA	992402	44747
Off2_264	2	female	CCCCA	842405	44255
Off2_265	2	male	AGTCA	1076346	45656
Off2_89	2	female	TTTTA	582371	41919
Off2_268	2	male	CGGCG	3635162	50707
Off2_215	2	male	GATCG	1664138	47843
Off2_219	2	male	GCATT	3709458	50865
Off2_274	2	male	GGTTC	1832027	47662
Off2_279	2	female	TATAC	1677729	49252
Off2_280	2	female	TTCCG	1522165	47000
Off2_281	2	female	AGCTG	1404143	47065
Off2_282	2	female	CTTCC	1391431	47148
Off2_283	2	female	CATGA	1172667	46450
Off2_286	2	female	GTACA	4529393	51462
Off2_287	2	male	ATTAG	741231	43416
Off2_288	2	male	CACAG	452378	39744
Off2_289	2	female	TCTCT	755258	44730
Off2_290	2	female	TGCAA	1432096	46707
Off2_292	2	female	GAAGC	3786041	53565
Off2_293	2	male	GGGGA	2091635	48616
Off2_294	2	female	TAATG	2969304	50281
Off2_296	2	male	TCAGA	2669023	48221
Off2_297	2	female	AACCC	2393732	48143
Off2_298	2	male	ACCAT	2026398	48851
Off2_300	2	male	ACTGC	3171799	49744
Off2_90	2	male	GGAAG	3062586	49758
Off2_147	2	female	TTGGC	2195986	48599