Sex Determination and Genetic Management in Nile Tilapia using Genomic Techniques

A Thesis Submitted for the Degree of

Doctor of Philosophy

By

Taslima Khanam

Institute of Aquaculture, Faculty of Natural Sciences,
University of Stirling, Stirling, UK

January 2017

Declaration

I declare hereby that this PhD thesis is entirely my own work and has been composed by myself. The works have been accomplished independently and all findings are original. Where the other sources of information have been used, this has been properly acknowledged and referenced. This work has not been currently or concomitantly and will not be submitted for any other degree.

Signature of Candidate:

Taslima Khanam

Signature of Principal Supervisor:

Prof. Brendan J. McAndrew

Signature of Supervisor:

Dr. David J. Penman

Date: January 2017

Acknowledgements

My PhD has been an amazing experience and it would have never been achievable without the generous support and guidance that I received from very important and special persons.

The first person I would like to express my heartfelt gratitude to my Principal supervisor Prof. Brendan J. McAndrew for accepting me as a PhD student under his scholastic supervision. I would like to say a very big thank you for all the supports and encouragements during the study period. Thank you very much for being there always and quickly sorted when I needed any support, which made my PhD a wonderful experience. I sincerely appreciate your active participation during thesis/manuscript write up and for valuable suggestions. I would like to express my indebtness to my supervisor Dr. David J. Penman wholeheartedly; you have been a tremendous mentor for me. Your kind support, intellectual stimulation, constant valued guidance, judicious constructive criticism and encouragement made my PhD a pleasing experience and enabled me to grow as a researcher. Thank you very much Dave for all the time invested in me to improve my research thinking and keeping your door always open to talk to you, and active participation while I was doing experiments and writing up articles/thesis. Your willingness to introduce me with the renowned scientists has been priceless. I am very pleased that I did my PhD research under both of your supervisions.

I feel short of vocabulary to put across my profound gratitude to Dr. John B. Taggart, who has been truly a dedicated mentor. Thanks very much John for sharing your expertise with me, your guidance during the lab work, fantastic hands-on training on NGS techniques and bioinformatics, and making me understand the technical part of the genetics and genomics field in a very friendly environment. Special thanks to Dr. John Benzie and Dr. Hugues de Verdal (WorldFish Center) for their excellent collaboration during my PhD period. Many thanks to Dr. Stefanie Wehner, it would have never been possible to analyse the NGS data without your expertise. Thanks to Dr. Michaël Bekaert (although you were there at the end of my PhD) to share your expertise with me and for your guidance during the data analysis. I would like to extend my sincere thanks to Dr. Jean-Francois Baroiller and Dr. Helena D'Cotta (CIRAD) for sharing their research experiences and for the discussions we had.

My special thanks to Mr. Keith Ranson, Mr. Brian Craig and Miss Kirsten McCallum, Dr. Kerry Bartie and Dr. Sarah-Louise Counter, your support was extremely helpful in the Tropical Aquarium Facilities and in the laboratory. Mr. Charlie Harrower, Miss Jane Lewis and Melanie Cruickshan were very helpful to provide all the research materials and for helping me out to go to different countries for Conference/Training.

I would like to thank past and present members of the Genetics research group Dr. Milton, Dr. Christos, Dr. Syaifudin, Khalfan, Dr. Munevver, Dr. Joanna, Sienna, Savitree for sharing the ideas and for their friendship. I would like to thank my friends Dr. Aliya, Dr. Olga, Dr. Fani, Aisha, Hayat, Rita, Houra, Munira, Ying, Suleiman and Sumaiya, Ahmed and Shomous, Angela, Yehwa, Margaret, Dr. Greta, Lucas, Michael and the list goes beyond……., whose company was always there. Many thanks to the Nepali family (Madhab Dada, Saru Didi, Mansi and Prisha) for their support and care. I would like to take the privilege to extend my acknowledgements to the Bengali community in the UK for their constant hospitality and making me feel at home always. I am grateful to Prof. Dr. Md. Abul Khair Chowdhury (BAU) for his endless encouragements, inspirations and supports; and many thanks to Prof. Dr. Mohammad Mahfujul Haque Ripon (BAU) for inspiring me to pursue my PhD at the Institute of Aquaculture.

My parents, I am so grateful to you for bringing me up with blessings, love, care, sacrifices, immense patience and inspirations (always with me). My heartfelt thanks to my lovely little sister (Lira), my beloved brothers (Nipu and Nithu), my uncle (Sanu mama) for their love, mental supports, blessings and wanted the best for me always. I would like to thank my mother-in-law and in-laws who supported me with blessings and care during my PhD.

My best friend and husband Dr. Atiqur Rahman, thank you very much for your unending and unconditional love and supports, inspiring me always to achieve my dreams, teaching me how to be patient, optimistic and think positive. Thanks for making my every moment enjoyable and being with me always in my success and also when I lost my way.

Finally special thanks to the Commonwealth Scholarship Commission (UK) for sponsoring me the PhD Scholarship and University of Stirling for hosting me as a PhD researcher and to have all the experiences mentioned above and beyond. Many thanks to the Genetic Society (GS), the Fisheries Society of the British Isles (FSBI) for providing grants to attend the international conferences, and the European Aquaculture Society (EAS) for their support in the Aquaculture Europe Conference 2016.

I humbly express my gratefulness to the Almighty and I sincerely believe that without His blessings, I would not have come all this way.

Taslima Khanam
January 2017

Dedicated

To

The departed soul of my father

Published Articles, Conferences and Trainings

List of Published Papers in Peer Reviewed Journals

- **Taslima K.**, Davie A., McAndrew B.J., Penman D.J. 2016. DNA sampling from mucus in the Nile tilapia, *Oreochromis niloticus*: minimally invasive sampling for aquaculture-related genetics research. Aquaculture Research, 47:4032-4037.
- **Taslima K.**, Taggart J.B., Wehner S., McAndrew B.J., Penman D.J. 2016. Suitability of DNA sampled from Nile tilapia skin mucus swabs as a template for ddRAD-based studies. Conservation Genetics Resources, doi: 10.1007/s12686-016-0614-z.

Conference Presentations

- **Taslima K.**, Khan M.G.Q., Taggart J.B., Wehner S., de Verdal H., Benzie J., McAndrew B.J., Penman D.J. 2016. Sex determination in Nile tilapia (*Oreochromis niloticus*) varies among populations. Aquaculture Europe, Edinburgh, Scotland, 20-23 September.
- **Taslima K.**, Taggart J.B., Wehner S., Andrew D., McAndrew B.J., Penman D.J. 2016. Minimally invasive DNA sampling from fish mucus for standard genotyping and next generation sequencing. Annual Symposium of the Fisheries Society of the British Isles-Fish, Genes & Genomes: Contributions to Ecology, Evolution & Management in Bangor University, North Wales UK, 18-22 July.
- **Taslima K.**, Taggart J.B., Wehner S., de Verdal H., Benzie J., McAndrew B.J., Penman D.J. 2016. Identifying the location of the major sex determining locus in Genetically Improved Farmed Tilapia using BSA-ddRAD sequencing. 4th International Symposium on Genomics in Aquaculture (GIA), Greece, 20-22 April.
- **Taslima K.**, Taggart J.B., Bekaert M., Penman D.J., McAndrew B.J. 2015. The use of RADseq for marker-assisted selection for controlling sex-ratio in Nile tilapia. 4th PhD Research Conference, Institute of Aquaculture, University of Stirling UK, 18th February.

Training Courses Attended

- An introduction to solving biological problems with Perl, University of Cambridge, UK, September 2016.
- An introductory R course, Institute of Aquaculture, University of Stirling, UK, November 2015.
- Introduction to RNAseq and ChIPseq data analysis, University of Cambridge, UK, August 2015.
- From genome-wide association studies (GWAS) to function, Italy, June 2015.
- Introduction to RADseq data analysis workshop, Edinburgh Genomics, University of Edinburgh, UK, May 2015.
- A general introduction to Minitab, Institute of Aquaculture, University of Stirling, UK, April 2015.
- Home Office animal handling modules 1-3 (Fish) under the Animals (Scientific Procedures) Act 1986, University of Glasgow, UK, October 2013.

Abstract

The PhD research studied two aspects in tilapia, firstly the analysis of sex determination in Nile tilapia (evidence of complex sex-determining systems) and secondly the genetic management of the tilapia species, using different genomic analysis approaches. This research started with the development of two techniques: minimally invasive DNA sampling from fish mucus, which was found to be suitable for standard genotyping and double-digest restriction-site associated DNA sequencing – ddRADseq; and pre-extraction pooling of tissue samples for ddRADseq (BSA-ddRADseq), which was found to be suitable for identifying a locus linked to a trait of interest (sex in this case). The first molecular evidence concerning the sex determination in genetically improved farmed tilapia (GIFT) was described using BSA-ddRADseq. Given the multiple stock origin of GIFT, surprisingly only a single locus (in linkage group 23) was found to be associated with the phenotypic sex across the population. The first evidence of LG23 influence on phenotypic sex in the Stirling population of Nile tilapia was also found. Different combinations of estrogen hormones and high temperature were tested for feminising Nile tilapia: a combined treatment of estrogen hormone and high temperature was found to be more efficient in feminising Nile tilapia than the estrogen alone. A set of species-diagnostic SNP markers were tested which were found to be suitable to distinguish pure species (*O. niloticus*, *O. mossambicus* and *O. aureus*), and these were used to analyse species contribution to GIFT and a selected tilapia hybrid strain. The results of the current research added novel information to our understanding of sex determination in Nile tilapia, which will be helpful in the development of marker-assisted selection in GIFT and other Nile tilapia strains towards the production of all male offspring. The methods developed also have broader applicability in genetic and genomics research.

Table of Contents

Contents Page

List of Tables

List of Figures

List of Abbreviations

AI	Aromatase Inhibitor	EB	Elution Buffer
ADB	Asian Development Bank	EE	Ethynyl Estradiol
AFLP	Amplified Fragment Length Polymorphism	ENA	European Nucleotide Archive
AMH	Anti-Müllerian Hormone	ERE	Estrogen Responsive Elements
AMHR2	Anti-Müllerian Hormone Receptor type II	EtBr	Ethidium Bromide
ANOVA	Analysis of variance	FAC	Freshwater Aquaculture Center
ASPA	Animals Scientific Procedures Act	FDR	False Discovery Rate
BAC	Bacterial Artificial Chromosome	FEM	Female-determining Gene
BFAR	Bureau of Fisheries and Aquatic Resources	FOXL2	Forkhead Box L2
BHA	Butylated Hydroxyanisole	FRET	Fluorescence Resonant Energy Transfer
BHT	Butylated Hydroxytoluene	FSHR	Follicle Stimulating Hormone Receptor
BPB	Bromophenol Blue	FTA	Flinders Technology Associates
BR	Broad Range	GIFT	Genetically Improved Farmed Tilapia
BSA	Bulk Segregant Analysis	GSD	Gonadal Sex Determination
BSD	Behavioural Sex Determination	GSDF	Gonadal Soma-derived Growth Factor
CLSU	Central Luzon State University	GtHRs	Gonadotropin Hormone Receptors
CM	Centi Morgan	HMG	High Mobility Group
CYP19A1	Cytochrome P450 family 19 subfamily A member 1	HotSHOT	Hot Sodium Hydroxide and Tris
ddRADseq	Double-Digest Restriction-site Associated DNA Sequencing	HPF	Hours Post Fertilisation
DES	Diethylstilbestrol	HS	High Sensitivity
DM	DNA-binding Motif	HT	Hybridization Buffer
DMRT	Doublesex and Mab-3 Related Transcription	ICLARM	Center for Living Aquatic Resources Management
DMW	W-linked DM-domain	IGF	Insulin-like Growth Factor
DMY	Y-linked DM-domain	IUCN	International Union for Conservation of Nature
DNA	Deoxyribonucleic Acid	KASP	Kompetitive Allele Specific PCR
dNTP	Deoxyribonucleotide Triphosphate	LG	Linkage Group
DPF	Days Post Fertilisation	MAS	Marker Assisted Selection
dsDNA	Double Stranded DNA	MID	Molecular Identifier
DSX	Doublesex		

MIH	Müllerian Inhibiting Hormone	SLS	Sample Loading Solution
MIS	Müllerian Inhibiting Substance	SNP	Single Nucleotide Polymorphism
NEB	New England Biolabs	SNV	Single Nucleotide Variant
NFFTRC	National Freshwater Fisheries Technology Research Center	SOX3	Sry-Box 3
NGS	Next Generation Sequencing	SRA	Sequence Read Archive
NT	Nucleotide	SRY	Sex-determining Region of the Y chromosome
NTC	No Template Control	SSR	Simple Sequence Repeats
PCR	Polymerase Chain Reaction	SXL	Sex Lethal
PGC	Primordial Germ Cell	TAF	Tropical Aquarium Facilities
PIT	Passive Integrated Transponder	TDF	Testis-Determining Factor
PPT	Parts Per Thousand	TGF-β	Transforming Growth Factor Beta
QTL	Quantitative Trait Loci	TRA	Transformer
RADseq	Restriction-site Associated DNA Sequencing	TSD	Temperature Sex Determination
RAPD	Random Amplified Polymorphic DNA	UNDP	United Nations Development Programme
RFLP	Restriction Fragment Length Polymorphism	UPMSI	Marine Science Institute of the University of the Philippines
RRL	Reduced Representation Library	WFC	WorldFish Center
RAS	Recirculating Aquaculture Systems	WGS	Genome-wide Sequencing
SAM	Sequence Alignment/Map	YSR	Yolk Sac Resorption
SCAR	Sequence Characterized Amplified Region	ZFAND3	Zinc Finger AN1-type Domain 3
SDY	Sexually Dimorphic an Immune-related gene on the Y-chromosome	ZFY	Y-linked Zinc Finger Protein

Structure of the Thesis

The PhD dissertation is divided into eight chapters, of which five chapters (chapter 3, 4, 5, 6, 7) are based on the experiments conducted during the research period.

The 1st chapter "general introduction" provides the background information used to design the present study and goes into detail about some aspects of the research questions for this study. **The 2nd chapter** describes "general materials and methods" including the general maintenance of the relevant biological materials, the experimental setup and the laboratory techniques. **The 3rd chapter** describes the development of DNA sampling technique from fish mucus and how the mucus-derived DNA has been used for standard genotyping and for double-digest restriction-site associated DNA (ddRAD) sequencing; and the results from mucus-derived DNA was compared with more invasive DNA sampling such as tissue biopsy. **The 4th chapter** describes the method development of pre-extraction pooling of tissue samples (BSA) for ddRADseq which would allow studying lots of individuals in a cost and time effective way. It also describes the mapping of the sex-determining region in GIFT across the population using this approach followed by the verification of the identified sex locus using different DNA-based markers. **The 5th chapter** describes the evidence of another sex-determining locus (LG23) in Stirling population of Nile tilapia given the occurrence of intra and inter-population variation in Nile tilapia. **The 6th chapter** explains the efficacy of combined treatment of estrogen hormone and high temperature for feminising Nile tilapia. Different batches of progenies were produced with different genetic sex combinations (using tightly sex-linked markers) and treated with different combinations of hormones and high temperature. **The 7th chapter** explains the use of species diagnostic markers (SNP) to determine the purity of the different tilapia species and to identify the genomic composition of the Molobicus hybrid tilapia (crosses from GIFT and feral *O. mossambicus*) after seventh generations of selection. **The 8th chapter** summarises the major findings, limitations of the current research and what is going on to the tilapia sex determination including future directions.

Chapter 1

General introduction

1.1 Sex determination and differentiation

Sex determination (SD), a fundamental step in the life of an organism, can be defined as the complex system of interacting biochemical processes that determine the sex of an individual (Hayes 1998). Sex differentiation is the follow-up process of sex determination where the subsequent development of distinct types of gonads, either ovary or testis, occurs once the decision has been made to become either male or female (Sandra and Norma 2010).

The discovery of sex chromosomes has driven researchers to uncover the complex systems of sex determination in a wide range of organisms. In 1891, H. Henking first reported a morphological feature of chromosomes during spermatogenesis from the hemipteran insect *Pyrrhocoris apterus*, where the unusual staining and behaviour of chromatin was observed during meiotic divisions. This chromatin body was dubbed as the X chromosome (Solari 1994). It has been found that the number of chromosomes is equal for both male and female in most animals. In exceptional cases, when chromosomes are unequally represented in the sexes, usually these chromosomal differences are related to sex determination.

1.2 Mechanisms of sex determination and differentiation

Sex is usually governed by three consecutive stages in almost all vertebrates. Generally sex is determined at the moment of fertilisation, which is called chromosomal sex determination. Once the decision has been made, the bipotential gonad differentiate into either an ovary or testis termed as gonadal sex differentiation. Finally, gonadal hormones regulate the secondary sexual characteristics expressed as either female or male phenotype. The mechanism of sex determination does not appear to be consistent from one group of animals to another, and varies even within closely related species and sometimes within

species too (Devlin and Nagahama 2002; Desjardins and Fernald 2009; Sandra and Norma 2010).

Sex determination mechanisms can be categorized into two broad classes: i) genotypic sex determination - GSD and ii) environmental sex determination - ESD (Volff *et al*. 2007; Penman and Piferrer 2008; Kobayashi *et al*. 2013). In GSD systems, the sex of an individual is directly determined by the genetic factors inherited from the parents, which decides whether the bipotential gonads differentiate into either ovary or testis without any external influences – e.g. human. Now a day, a gradient of GSD types is considered, between differentiated chromosomes, through single gene(s) without differentiated sex chromosomes, "oligogenic" systems (e.g. Nile tilapia) to genuinely polygenic systems (e.g. European seabass). In some species, environmental factors such as temperature, pH, salinity, photoperiod and social behaviour sometimes play significant roles in sex determination and gonad differentiation (Devlin and Nagahama 2002; Baroiller *et al*. 2009a). Effects of temperature on sex determination and differentiation in animals (temperature sex determination, TSD) has been studied widely and reported to have a strong influence on sex in certain groups of insects, many turtles, some lizards, all crocodiles and some fishes where the incubation temperature of the embryo directs the sex. Some authors consider that genuine TSD is adaptive and seen in the species in the wild, rather than the artefactual effects we see in species such as seabass (probably only happens when the temperature is raised in the hatchery). Reptiles were the first animals reported to have TSD (Ospina-Alvarez and Piferrer 2008). Sometimes a combination of both genetic and environmental factors (GSD + ESD) also determines the sex of an individual such as European sea bass, *Dicentrarchus labrax* (Vandeputte *et al*. 2007).

1.3 Sex determination in human and other animals

In mammals, sex is determined at the moment of fertilisation by differential inheritance of sex chromosomes where an XX embryo becomes female and an XY embryo is destined to be a male. Before the discovery of human sex chromosomes (X and Y), it was thought that sex was determined by the ratio of the number of X chromosome and the number of sets of autosomes - X:A (Painter 1923). Later on, it was found that the Y chromosome controls mammalian sex determination. During the late 1950s and early 1960s, XO mice and XO humans were reported without testicular tissue and it was postulated that mammalian sex depends on the Y chromosome (Ford *et al*. 1959). The development of XXY humans and XXY mice as phenotypic males with testis further proved that testis development depends on presence or absence of the Y chromosome (Jacobs and Strong 1959).

All therian mammals, both placental and marsupial have a male heterogametic (XX/XY) sex chromosome system where dosage and time-dependent action of a series of sex-related genes determine the sex (Huyhn *et al*. 2011; Jakob and Lovell-Badge 2011; Angelopoulou *et al*. 2012). Mammalian sex is first determined genetically followed by hormonal regulation of the developing phenotype. The genetic pathway is controlled by a master switch, the testis-determining factor (TDF), that initiates the formation of a testis in XY and anti-Mullerian hormone (*Amh*) that signals the sexual differentiation in males. In the absence of a master switch, the undifferentiated gonad transformed into an ovary (Waters *et al*. 2007). A Y-linked zinc-finger protein gene (*Zfy*) was first thought to be TDF (Page *et al*. 1987) but later mutation analysis of an XY female indicated that the *Sry* (sex-determining region Y) gene is responsible for sex determination in mammal (Sinclair *et al*. 1990).

Sry evolved from the closely related *Sox3* gene (High Mobility Group - HMG box DNA binding domain), which is on the X chromosome and has no relation with sex determination

(Cutting *et al*. 2013). In the few placental mammalian species, which do not have *Sry*, such as mole vole (*Ellobius lutescens)* and Zaisan mole vole (*E. tancrei*), the XO and XX sex chromosomes are identical for both male and female. The spiny rat, *Tokudaia simensis*, also has an XO complement in both sexes (Arakawa *et al*. 2002). These rodents have different systems of sex determination without the involvement of the Y chromosome or *Sry*, the mechanism of which is yet to be determined (Ferguson-Smith 2007).

In the XY gonad of human and nearly all other placental mammals, the Y-linked *Sry* gene with some positive regulators trigger the upregulation of different downstream genes which contribute to the development of a functional testis (Jakob and Lovell-Badge 2011). Whereas in the XX gonad, negative regulators supress testis gene expression by promoting ovarian gene expression (Uller and Helanterä 2011).

Monotremes including Platypus and Echidna have very complicated sex chromosome systems and form complicated linked structures in meiosis. The Platypus female has five pairs of X chromosomes ((X1X1 X2X2 X3X3 X4X4 X5X5) and the male has 10 unpaired chromosomes (X1Y1 X2Y2 X3Y3 X4Y4 X5Y5). The sex chromosomes of Echidna share homology with the Platypus but they have five X and four Y chromosomes. There is no homology of Platypus X chromosomes with the therian X chromosomes (Veyrunes *et al*. 2008) but the X5 chromosome is highly homologous with the Z chromosome of bird (Rens *et al*. 2007).

In birds, sex is determined genotypically at the moment of fertilisation and all species have a WZ/ZZ sex chromosome system where the female is the heterogametic - WZ and male is homogametic - ZZ (Smith *et al*. 2007). Sex in birds is determined by the dominant female determining gene on the W chromosome or dose dependent male determinant on the Z chromosome (Ellegren 2001; Smith *et al*. 2007). As birds lack *Sry* (Chue and Smith 2011),

the best candidate gene (Doublesex and mab-3 related transcription factor1 - *Dmrt1*) for male sex determination is located on the Z chromosome (double gene dosage) and absent from the W chromosome in all birds (Shetty *et al*. 2002; Smith *et al*. 2003; Ezaz *et al*. 2006; Siegfried 2010). It has been experimentally proven that *Dmrt1* is a master regulator of testis differentiation in birds (Zhao *et al*. 2010). Although it has been postulated that the genetic factors are responsible for sex determination in all birds, there is evidence that the sex of some birds is also influenced by environmental factors. For example incubation temperature of the eggs affects sex-ratio in the Australian brush-turkey megapode, *Alectura lathami* where lower incubation temperatures trigger male and higher incubation temperatures triggers female skewed sex-ratio (Göth and Booth 2005).

Reptiles exhibit an impressive and confusing array of sex chromosome structure and sex determination mechanisms extending from genotypic to environment (Valenzuela and Lance 2004). All snakes exhibit genotypic sex determination in which sex is chromosomally determined at the moment of fertilisation, whereas all crocodile and marine turtles exhibit environmental sex determination, especially incubation temperature of the eggs drives the sex-ratio (Schartl 2004; Ferguson-Smith 2007; Hawkes *et al*. 2009). Turtles and lizards exhibit both genotypic and environmental sex determination. Female heterogamety (WZ, WZZ, or WWZ) is observed in snakes whereas turtles show male heterogamety (XY or XXY), and both male and female heterogamety are observed in lizards (Sites *et al*. 1979; Sarre *et al*. 2004). The first evidence of temperature sex determination was observed in lizards (*Agama agama*) almost 50 years ago (Janzen and Phillips 2006; Smith *et al*. 2007). Mostly temperature of the nest environment and sometimes weight and size of eggs determines whether an egg develops as either male or female (Pandey and Nandi 2014**).** High temperature triggers sex-ratio towards male development in crocodiles whereas in turtles higher temperature promotes female

development (Schartl 2004; Wright *et al*. 2012). It is possible to produce sex reversed lizards by altering the incubation temperatures or by the application of steroid to the eggshell (Radder *et al*. 2008). Species with TSD (sometimes GSD) do not display karyotypic differences between males and females and very little differences ($1^{o}C$) in temperature can produce all male or all female individuals (Crews *et al*. 1994; Modi and Crews 2005).

All species of amphibians studied to date exhibit genotypic sex determination of either male - XX/XY or female heterogamety - WZ/ZZ or both (Schmid and Steinlein 2001; Nakamura 2009; Schmid *et al*. 2010). For examples the Japanese wrinkled frog, *Rana rugosa* has a complex sex-determining system; heteromorphic XY, homomorphic XY and heteromorphic WZ sex chromosomes within a single species (Nishioka *et al*. 1993; Schartl 2004). Recently a candidate gene *Dmw* (W-linked DM-domain gene) has been identified from African clawed frog, *Xenopus laevis* and is thought to be involved in ovary determination (Yoshimoto *et al*. 2008; Yoshimoto *et al*. 2010).

From the very beginning, insects were the center of genetic research interest due to their diversified worldwide distribution. The first evidence of a sex-determining mechanism and inherited basis of sex determination was reported from insects (McClung 1902; Wilson 1905), and insects exhibit an astonishing diversity of sex-determining systems. Both male heterogametic (XX: female; XY: male) and female heterogametic (ZZ: male; WZ: female) systems are evident in some species and some species do not have any sex chromosomes (Verhulst *et al*. 2010). Sex of insects is also determined at the ploidy level and most species are diploid (2n) but some species including wasps, ants, thrips and mites are haplo/diploid - males: n, females: 2n (Kageyama *et al*. 2012).

In the haplodiploid sex-determining system, males are haploid developed from unfertilised eggs via arrhenotokous parthenogenesis, receiving only a single set of maternal chromosomes, whereas females are diploid developed from fertilised eggs and inherit both maternal and paternal chromosomes. Female heterogametic sex-determining systems (WZ/ZZ and Z/ZZ) have been found in species-rich insect orders like Lepidoptera (moths and butterflies) and their closest relatives, Trichoptera (caddis flies). A female determining gene (*Fem*), present on the W chromosome, determines their sex. Sporadic rearrangements created multi-sex chromosome systems (W_1W_2Z/ZZ and $WZ_1Z_2/Z_1Z_1Z_2Z_2$) and sporadic losses of the W chromosome changed the system formally back to Z/ZZ in some species (Traut *et al*. 2007; Suzuki 2010).

In *Drosophila melanogaster*, a well understood model species, sex is determined independently by the ratio of the number of X chromosomes and the number of sets of autosomes- X:A (Cline 1993). When the ratio (XX:AA) is 1, the developing insect is female, whereas a ratio (XY:AA) of 0.5 develops as male (Saccone *et al*. 2002; Vicoso and Bachtrog 2013). Once sex is determined, four main genes (Sex-lethal, *Sxl*; transformer, *Tra*; transformer-2, *Tra-2*; and doublesex, *Dsx*) together maintain sex differentiation during later developmental stages (Kageyama *et al*. 2012).

1.4 Sex determination and differentiation in fish

1.4.1 Sex determination

Fishes exhibit a diverse array of reproductive strategies and are categorized into gonochorism, synchronous/sequential hermaphroditism or unisexualism (Atz 1964). Sex determination and differentiation systems in fish are more flexible and plastic because of the relatively common occurrence of hermaphroditism and naturally occurring sex-reversal.

The most common mode of sex determination in fish is genotypic sex determination (GSD), where sex is determined by both major (chromosomal systems) and minor (polyfactorial) genetic factors. Chromosomal sex determination can be male heterogametic - XX/XY; female heterogametic - WZ/ZZ and sometimes involve several pairs of sex chromosomes (X1X1X2X2/X1X2Y, W1W2Z/ZZ). In some cases, three or more major sex-determining factors contribute to sex development (Platyfish - *Xiphophorus maculatus* where male can be XY or YY and female can be XX, WX or WY). In gonochoristic fish species (develop either as male or female and maintain their sexual status throughout their life), almost all possible kinds of sex-determining systems have been observed (Devlin and Nagahama 2002; Schartl 2004).

There are plenty of methods to elucidate the mechanism of sex determination in fish species. One of the quickest methods is to analyze the karyotype of male and female individuals to identify genotypic sex determination involving differentiated sex chromosomes. But many fish species do not exhibit heteromorphic sex chromosomes; only 32 out of about 900 neotropical freshwater fish species display morphologically differentiated sex chromosome (de Almeida Toledo and Foresti 2001). The first evidence of heteromorphic sex chromosomes was found in two unnamed Mexican fish, which had a clearly visible Y chromosome in males (Uyeno and Miller 1971, 1972). The advent of DNA technologies has opened up new windows to identify sex-linked and sex-specific markers, and key regulators or genes responsible for sex determination in a variety of fish species. The sex-determining region of some fishes has already been discovered such as Chinook salmon, *Oncorhynchus tshawytscha* (Devlin *et al*. 1991; Stein *et al*. 2001); Channel catfish, *Ictalurus punctatus* (Liu *et al*. 1996); Medaka, *Oryzias latipes* (Matsuda *et al*. 2002; Kondo *et al*. 2006); Platyfish, *X. maculeatus* (Froschauer *et al*. 2002); Nile tilapia, *O. niloticus* (Lee *et al*. 2003); Threespine stickleback, *Gasterosteus aculeatus* (Peichel *et al*. 2004); Rainbow

trout, *Oncorhynchus mykiss* (Felip *et al*. 2005) and Tiger pufferfish, *Takifugu rubripes* (Kikuchi *et al*. 2007).

Environmental sex-determining mechanisms are associated with abiotic factors such as temperature, pH, salinity, photoperiod, density, hypoxia and social behaviour. Temperature is one of the most important factors and has been studied extensively in fish sex determination (TSD). Environmental sex determination sometimes can override the genetic sex determination by directly influencing the genome via the neuroendocrine system or through direct alterations of hormone production within the fish body (Aylmer 1930; Penman and Piferrer 2008). Effects of environmental factors on skewed sex-ratio may differ among different strains of the same species based on their genetic constitution and relative strength of the sex factors (Mylonas *et al*. 2005). Temperature triggers the skewed sex-ratio in three different ways depending on the species: 1) in the majority of species, exposure to high temperature increases number of males and a female skewed ratio is induced by lowering the temperature, 2) in a few species like channel catfish, *Ictalurus punctatus*, female percentage increases with increasing temperature while low temperature induces higher male percentages and 3) at extreme temperatures (high and low), male skewed sex-ratios are produced while an intermediate temperature generates a 1:1 sex-ratio, e.g. Southern flounder - *Paralichthys lethostigma* (Luckenbach *et al*. 2003; Ospina-Alvarez and Piferrer 2008; Luckenbach *et al*. 2009). The first evidence of temperature sex determination (TSD) in fish was reported in Atlantic silverside, *Menidia menidia,* by Conover and Kynard (1981) where high temperature during the critical period of larval development resulted in a higher proportion of males and this is "real" TSD, i.e. its ecological significance is understood; for many other species, the temperature effects may be artefactual, i.e. observed in experimental/hatchery conditions only, and not experienced (or only very rarely) by the species in the wild. TSD has been studied on more than 60 fish species, of which tilapia

(Baroiller *et al*. 1995); pejerrey, *Odontesthes bonariensis* (Struussmann *et al*. 1996); channel catfish, *Ictalurus punctatus* (Patiño *et al*. 1996); sockeye salmon, *Oncorhynchus nerka* (Craig *et al*. 1996); Japanese hirame, *Paralichthys olivaceus* (Yamamoto 1999); goldfish, *Carassius auratus* (Baroiller *et al*. 1999) are some examples. Environmental factors other than temperature were also studied including the effect of pH on *Apistogramma* (Römer and Beisenherz 1996); photoperiod on European sea bass, *D. labrax* (Bláquez *et al*. 1998); salinity on Nile tilapia, *O. niloticus* (Abucay *et al*. 1999) and European sea bass, *D. labrax* (Saillant *et al*. 2003) and stocking density on Paradise fish, *Macropodus opercularis* (Francis 1984). Behavioural sex determination (BSD) was first observed in coral reef fishes in the early 1970s where they can change their sex in the absence of the opposite sex (Fishelson 1970).

In almost all gonochoristic species, the undifferentiated gonads directly develop into either ovaries or testes. In some gonochoristic species gonads initially develop as ovaries and after a certain period, the ovaries degenerate and develop into a testis (*Danio rerio* and *Barbus tetrazona*). Hermaphroditism in fishes is widespread; about 2 % of all teleost fish species from 27 families have shown sequential hermaphroditism. Protogyny and protandry are the two main types of sequential hermaphroditism. Some fish first mature as female and eventually change their sexual status to become functional males; these are known as protogynous hermaphrodites, for example *Lates calcarifer*, *Thalassoma duperrey* (Hourigan *et al*. 1991). At the initiation of sex change, vitellogenic oocytes degenerate and there is a rapid drop in plasma estradiol-17β followed by a gradual increase in 11-keto testosterone which increases the formation of testicular tissues. In early stages of sex change, *Dmrt1* and gonadal soma derived factor (*GSDF*) are the key regulators for testicular differentiation in other vertebrates and also involved in the development of spermatogonia (Kobayashi *et al*. 2013). In protandrous hermaphroditism, fish mature as male and then change their sex to

female, for example *Amphiprion* sp., *Sparus aurata* (Godwin *et al*. 1996). Sex steroid especially estrogen plays a vital role in changing the sex of coral reef fish from male to female (Nakamura *et al*. 2005; Kobayashi *et al*. 2010). Bi-directional sex change (male to female and female to male) is another type of hermaphroditism (reversible), and Okinawa rubble gobiid (*Trimma okinawae*) is the first known bi-directional sex change hermaphrodite (Sunobe and Nakazono 1993). Social visual cues are responsible for initiation of this sex change followed by changes of gonadotropin hormone receptors (GtHRs) in the gonad (Kobayashi *et al*. 2013).

1.4.2 Sex differentiation

The process of fish sex differentiation is becoming better understood, at least from a morphological and endocrinological point of view (Devlin and Nagahama 2002). In the process of sex differentiation, from the early stage of germ cell migration and gonadal ridge formation to the final stages of gonad formation, many genes are involved. Moreover, many of these genes are common to all vertebrates, despite the fact that extremely diverse determinants trigger the sex determination in non-mammalian vertebrates (Place and Lance 2004). Genes downstream of the master regulatory gene that establish sexual dimorphism during gonadal development and the associated mechanisms are quite conserved in fish (Piferrer and Guiguen 2008).

1.4.3 Key regulators for sex determination and differentiation

A gene regulatory cascade triggers the development of sex in both invertebrate and vertebrate animals. The first vertebrate sex-determining genes to be identified where *Sry* in mammals and *Dmy* in the teleost fish Medaka (*O. latipes*) (also designated as *Dmrt1by* is a duplicated copy of the autosomal gene *Dmrt1*), both genes are linked to the Y chromosome

(Koopman 2001; Matsuda *et al*. 2002; Nanda *et al*. 2002; Siegfried 2010). Birds have an WZ/ZZ sex-determining system, where the male is homogametic ZZ and female is heterogametic WZ. The Z linked *Dmrt1* gene (absent in W chromosome) is a master regulator for testis formation leading to male sex development in birds (Nanda *et al*. 1999; Smith *et al*. 2009).

Doublesex and Mab3-related transcription factor, *Dmrt*, was first identified in fruit fly (*Drosophila melanogaster*) termed as *Dsx* (doublesex, Baker and Wolfner 1988) and in roundworm (*Caenorhabditis elegans*) as *mab*-3 (Shen and Hodgkin 1988; Fernandino *et al*. 2008). The *Dmrt* gene family is characterized by intertwined zinc finger cysteine rich DNA binding motif termed the DM domain (Murphy *et al*. 2010; Kopp 2012). *Dmrt* gene is considered as downstream regulator in male sex determination and differentiation and has been found in phylogenetically divergent animals like mammals, birds, reptiles, amphibians, fish, flies, nematodes and corals (Herpin and Schartl 2011).

Homologues of *Dmrt* have been identified in different species linked to the sex chromosome, such as Y-linked *Dmy* (also called *Dmrt1by*) in the teleost fish Medaka, W-linked *Dmw* in the African clawed frog (*Xenopus laevis*) and Z-linked *Dmrt1* in the bird (Mawaribuchi *et al*. 2012). *X. laevis* has dominant female determining function where W-linked *Dmw* regulates female sex determination by repressing the transcriptional targets of *Dmrt1* (male determining gene) and upregulating the expression of WZ gonad-specific genes such as *Foxl2* and aromatase gene *Cyp19* during and after sex determination (Okada *et al*. 2009; Yoshimoto *et al*. 2010).

Eight DM domain genes have been identified to date in vertebrates namely *Dmrt1*, *Dmrt2a*, *Dmrt2b*, *Dmrt3*, *Dmrt4, Dmrt5*, *Dmrt6*, *Dmrt7* and *Dmrt8*. The latter two are restricted to mammal, *Dmrt6* is only found in tetrapods and *Dmrt2b* is found only in teleosts (Veith *et*

al. 2006; Hong *et al*. 2007). Male restricted *Dmrt1* expression has been reported in various gonochoristic fish species such as Medaka, *O. latipes* (Kobayashi *et al*. 2004); Olive flounder, *Paralichthys olivaceus* (Jo *et al*. 2007); Rare minnow, *Gobiocypris rarus* (Zhang *et al*. 2008); Nile tilapia, *O. niloticus* (Kobayashi *et al*. 2008) and North African catfish, *Clarias gariepinus* (Raghuveer and Senthilkumaran 2009). In some fishes *Dmrt1* has been expressed in testis and ovary with strong male biased expression, for example in Rainbow trout, *O. mykiss* (Marchand *et al*. 2000); Zebrafish, *Danio rerio* (Guo *et al*. 2005); Shovelnose sturgeon, *Scaphirhynchus platorynchus* (Amberg *et al*. 2010); Pejerrey, *Odontesthes bonariensis* (Fernandino *et al*. 2008); Atlantic cod, *Gadus morhua* (Johnsen *et al*. 2010); Lake sturgeon, *Acipenser fulvescens* (Hale *et al*. 2010) and Southern catfish, *Silurus meridionals* (Liu *et al*. 2010).

Sox9, the *Sry*-related high mobility group (HMG) containing box gene, has been involved in male determination in vertebrates (Siegfried 2010). The *Sox9* transcription factor is a DNA-binding protein that contains a 79 amino acid long HMG domain with at least 50 % similarity to that of *Sry* in mammal (Wright *et al*. 1993). In mammals, *Sox9* has been found to be a direct target of *Sry* and upregulates the expression of Anti-Müllerian hormone (*Amh*) gene (Sekido and Lovell-Badge 2009; Sekido 2010). But in non-mammalian vertebrates, the role of *Sox9* is not well understood (Siegfried 2010). The expression of *Sox9* gene has been reported in several teleosts including Zebrafish, Threespine stickleback, Fugu, Medaka, Rainbow trout, Nile tilapia and Common carp (Chiang *et al*. 2001; Cresko *et al*. 2003; Zhou *et al*. 2003; Koopman *et al*. 2004; Nakamoto *et al*. 2005; Vizziano *et al*. 2007; Du *et al*. 2007; Ijiri *et al*. 2008).

Amh also known as Müllerian inhibiting substance or factor (*Mis*) or Müllerian inhibiting hormone (*Mih*) was first proposed by Alfred Jost in 1940s (Rey *et al*. 2003). The *Amh* -

homodimeric disulphide-linked glycoprotein, is a member of the transforming growth factor-β (TGF-β) superfamily (Sandra and Norma 2010; Siegfried 2010). In mammal, bipotential gonads have two ducts such as the Müllerian duct (ovary differentiation) and the Wolffian duct (testis differentiation). *Sry* induces the bipotential gonad to differentiate into testis that produces two hormones, *Amh* and testosterone (Jamin *et al*. 2003). *Amh* expressed in both testis and ovary causes the regression of Müllerian ducts in the male fetus (Vigier *et al*. 1989; Josso 2011) and in females differentiate into fallopian tubules and uterus (Rodríguez-Marí *et al*. 2005). It is produced by Sertoli cells from the time of differentiation until puberty and in ovaries by granulosa cells from birth until menopause (Al-Attar *et al*. 1997). Despite the absence of Müllerian ducts, teleost still show Sertoli cell expression of *Amh* in fish species such as Japanese flounder (Yoshinaga *et al*. 2004), Zebrafish (von Hofsten *et al*. 2005), Nile tilapia (Ijiri *et al*. 2008), Rainbow trout (Vizziano *et al*. 2008) and Atlantic Cod (Johnsen *et al*. 2013). Y-linked *Amh* duplicated copy has been identified in fish species such as the Patagonian pejerrey (*Odontesthes hatcheri*, Hattori *et al*. 2012) and in Nile tilapia (Eshel *et al*. 2014; Li *et al*. 2015) and a receptor of *Amh* type II (*Amhr2*) has been identified in three different Takifugu species (*T. rubripes, T. pardalis* and *T. poecilonotus*; Kamiya *et al*. 2012).

Steroids are the major products of all vertebrate gonads produced by the activity of different enzymes (Piferrer and Guiguen 2008). The gene *Cyp19a1* encodes for the enzyme cytochrome P450 aromatase (P450arom) which catalyses the conversion of androgens to estrogens and is of great importance to sexual differentiation of many vertebrates including fish (Siegfried 2010). In both mammal and fish, *Foxl2* transcription factor has been known to trigger the transcription of *Cyp19a* gene (Wang *et al*. 2007b). Tetrapods have single copy of this gene whereas teleosts have two isoforms of the aromatase gene, *Cyp19a1a* (also called ovarian aromatase or gonadal aromatase) mainly expressed in the female gonads and

Cyp19a1b (also known as neural aromatase or brain aromatase) predominantly expressed in the brain (Kwon *et al*. 2002; Chang *et al*. 2005).

Another transcription factor *Foxl2* (forkhead box L2), is a member of the winged helix/forkhead group, which is also involved in vertebrate sex determination and differentiation (Piferrer and Guiguen 2008); and some fishes have two paralogues of *Foxl2* gene (Sandra and Norma 2010). A few more sex-determining genes have been identified in fish species such as *Gsdfy* in *O. luzonensis* (Myosho *et al*. 2012), *Sox3y* in *Oryzias dancena* (Takehana *et al*. 2014), *Sdy* in the Rainbow trout (*O. mykiss*) and other salmonids (Yano *et al*. 2012, 2013). The pathways and the regulatory factors involved in fish sex determination and differentiation are described in **Figure 1.1**.

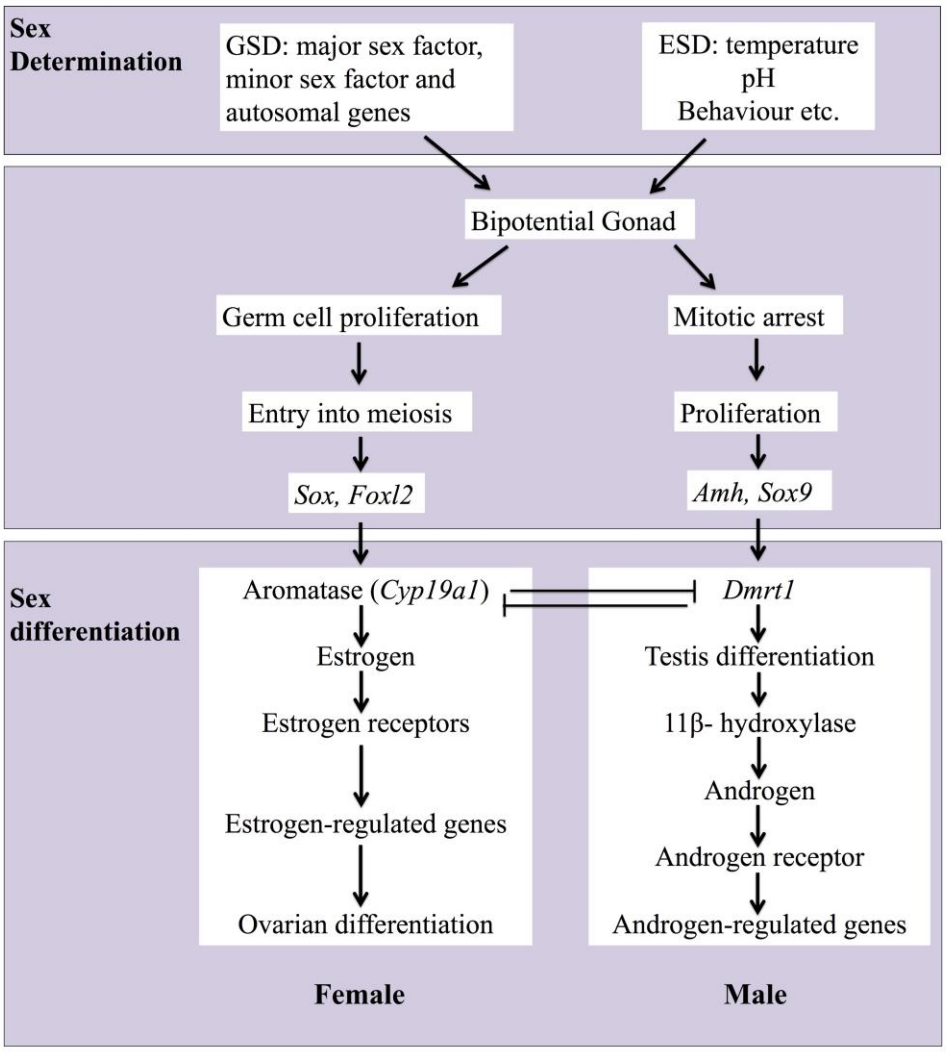


Figure 1.1 Schematic diagram of fish sex determination and differentiation. The figure is not representative of any particular fish, but presents a common pathway on what is currently known. *Cyp19a1* and *Dmrt1* are the main regulators of sex differentiation in fish; help to differentiate into ovary and testis respectively. The gonad developmental time is independent for fish species. (Original image, Piferrer and Guiguen 2008).

1.5 Biology of tilapia and culture potential

1.5.1 Background and general biology of tilapia

Fishes are an immensely diversified group of aquatic vertebrates. Teleosts comprise of about 26,000 species, which is almost half of all living vertebrates (Volff 2005). Among the

62 orders of teleost, Perciformes is the largest one, representing more than 10,000 species (Nelson 2006). The Cichlidae, belonging to the order Perciformes, is one of the most species-rich families of vertebrates, consisting of approximately 3,000 species (Volff 2005; Nelson 2006). The Cichlidae is a monophyletic group, divided into four sub-families namely Etroplinae, Ptychochrominae, Cichlinae and Pseudocrenilabrinae. Hemichromine, haplochromine and tilapiine are the three major tribes of Pseudocrenilabrinae cichlids (Ferreira *et al*. 2010).

The tilapiine tribe, commonly known as Tilapia, shows species diversity with more than 70 freshwater species and a few brackish water species belonging to the genera *Tilapia*, *Sarotherodon, Oreochromis* and *Danakilia* (Trewavas 1983; Penman and McAndrew 2000). The first three of the four genera are commercially the most important ones. *Tilapia* are biparental caring substrate spawners and *Sarotherodon* are generally paternal or biparental mouth brooders, while *Oreochromis* are only maternal mouth brooders. Nile tilapia, *O. niloticus*, plays a significant role in commercial aquaculture and accounts for more than 90 % of all commercially farmed tilapia globally. Additionally, Blue tilapia (*O. aureus*), Mozambique tilapia (*O. mossambicus*), Galilee tilapia (*Sarotherodon galilaeus*), Zanzibar tilapia (*O. urolepis hornorum*), Sabaki tilapia (*O. spilurus*), Black-chinned tilapia (*Sarotherodon melanotheron*), Congo tilapia (*Tilapia rendalli*), various hybrids mainly *O. niloticus* × *O. aureus* (China, Israel) and Red hybrid tilapia and Redbelly tilapia (*T. zillii*) also have some commercial importance in the international market.

Tilapias are the second most important aquaculture species in the world after carps. Although tilapia farming is widespread from Africa to nearly 140 countries elsewhere in the world, the majority of tilapia production comes from developing countries in Asia and Latin America. Asia leads the world in total tilapia production, and China is the top producer and

also exporter of tilapia in the world followed by Egypt, Indonesia, Philippines, Thailand, and Brazil.

Tilapias show superiority for aquaculture over many other fishes because of their ability to adapt to a wide variety of physico-climatic cultural conditions including poor water quality and environmental fluctuations and a wide range of aquatic environments from freshwater, brackish water, seawater and even in hypersaline water. They can tolerate wide variations in water temperature (8 to 43^{o}C), oxygen concentration (0.3 to 1.5 ppm), pH (6 to 10), salinity (0 to $\leq$ 37 ppt) and ammonia concentration (0.2 to 3 ppm) (Baroiller *et al*. 2009a). They grow very well, are disease resistant, have a short generation time and are easy to breed in captivity all the year round providing the temperature is high enough.

1.5.2 Genetic management of tilapia

Tilapias are among the leading aquaculture species in the world to meet the protein demand of the increasing human population. Their versatility in different aquaculture systems from extensive to highly intensive, and their adaptability and capacity to endure environmental fluctuation without adverse effect, has seen the tilapia farming industry continuing grow in Asia and other regions to fulfil the farmers' and consumers' demands. Among all tilapias, Nile tilapia (*O. niloticus*) has earned the greatest popularity among the farmers and consumers because of their good taste, colour, rapid growth, hardiness, resistance to disease, and ease of culture in most water environments except full strength seawater (McAndrew 2000).

The maintenance of genetic quality of the pure species and proper broodstock management are important for performance in aquaculture and the development of successful selective breeding, given the evidence of hybridization and gene introgression in the wild and captive

environments. The introduction of fishes from one place to another or the release of stocks to the wild are possible ways of introducing new gene into the existing species which in turn reduces the genetic purity of the species which may cause major declines in productivity. For example Mozambique tilapia (*O. mossambicus*) was the first introduced tilapia species in Asia. Introduction of the small number of fish in Asia (founder and bottleneck effects) caused a rapid decrease in production of Mozambique tilapia due to severe reduction in growth performance (Pullin and Capili 1988). Mozambique tilapia is also not a good aquaculture species (early maturation), so they were replaced by Nile tilapia. According to IUCN, *O. mossambicus* has now been categorized a "near threatened" in the wild in Africa because of its hybridization with introduced *O. niloticus* (Cambray and Swartz 2007). Asian stocks of introduced *O. niloticus* have been found to be introgressed with *O. mossambicus* due to poor broodstock management, which is the major cause of its declining genetic status (Taniguchi *et al*. 1985; Macaranas *et al*. 1986; Eknath *et al*. 1991). *O. esculentus,* was native to Lake Victoria and following the introduction of *O. niloticus* into the Lake Victoria, hybrid (*O. esculentus* and *O. niloticus*) have been detected, and after that there was no pure stock of *O. esculentus* remaining in the lake (Mwanja and Kaufman 1995). It has disappeared from the lake and now categorized as "critically endangered" species in the IUCN Red list (Twongo *et al*. 2006). With many different species of tilapia and many breeding programmes being initiated around the world, it is now an urgent need to identify large numbers of species-specific diagnostic markers to differentiate between species and hybrids. Protein and DNA-based molecular markers have been developed to differentiate tilapia species. None of the techniques has the ability to accurately differentiation among species, and estimate hybridization and gene introgression. Next generation sequencing (NGS) technologies offer great potential to generate large number of informative single nucleotide polymorphic (SNP) markers in a cost effective way (Mardis

2008a). Using these techniques, it would be possible to generate large number of single nucleotide polymorphic (SNP) markers, which will be able to distinguish different tilapia species ((Syaifudin 2015).

1.5.2.1 Non-invasive DNA sampling for genetic management of fish

Minimally invasive or non-invasive DNA sampling offers great potential for conservation and management applications in animal biology especially for endangered species. Good quality DNA is essential of studying genetic variation within and between population, and identifying evolutionary ancestries. There are various types of methods used for DNA sampling ranging from invasive to non-invasive. Non-destructive sampling methods e.g. fin, scale, gill, barbel, muscle, blood, sperms and eggs are mainly used as DNA sources for fish species (Campanella and Smalley 2006). Among them tissue biopsy is a widely used sampling method without considering its negative impact on fish. This may affect survival, growth, sometimes can alter the behaviour of fish and have the chance of secondary infection (Le Vin *et al.* 2011). On the other hand, blood sampling from fish is also practiced to extract DNA but it is difficult to sample blood from small fishes without sacrificing the fish. Blood sampling without sacrificing also requires trained personnel, involves anaesthesia and mishandling of fish sometimes cause problems to the fish (Pirhonen and Schreck 2003). Non-invasive or least invasive DNA sampling (either brush or FTA - Flinders Technology Associates, card) could be an effective way for fish, minimizing the adverse effects on fish following invasive DNA sampling. The non-invasive DNA sampling technique has been used for RFLP (Livia *et al*. 2006), RAPD and mtDNA (Hoolihan *et al*. 2009) and microsatellite (Le Vin *et al*. 2011) studies, which proved to be simple, reliable, relatively non-invasive, and to yield DNA suitable for such techniques. Next generation sequencing (NGS) technologies have the power to generate thousands of informative

markers in a cost-effective way, applicable to quantitative, structural, functional and evolutionary studies in model and non-model organisms (Metzker 2010). It is important to use DNA derived from non-invasive sampling to the NGS platform to increase the applicability of both NGS and non-invasive DNA sampling which will be useful for genetic management and conservation of vulnerable species. DNA derived from minimally invasive sampling for NGS has only been used on a limited scale, e.g. in humans (Ogawa *et al*. 2013) and birds (Vo and Jedlicka 2014), but has yet to be explored for fish.

1.5.3 Breeding programmes in aquaculture especially for tilapias

Aquaculture, a major source of animal protein, is the fastest growing food production sector, and Asia is the greatest contributor to the global aquaculture production. Selective breeding programmes have been successfully used for the genetic improvement of plants and terrestrial animals. It has been noticed that 6-8 fold higher genetic gain can be found from fish through selective breeding compared with what is found in the farm animal breeding (Gjedrem 2010), however selective breeding in aquaculture is far behind land animals. About 12.5 % genetic gain in terms of the growth per generation can be obtained from fish which means the aquaculture production can be dramatically increased if the genetically improved farmed fish are used. It has been estimated that less than 10 % of the aquaculture production comes from the genetically improved stocks despite the fact that genetic gain for aquaculture species is substantially higher than the terrestrial animal (Gjedrem *et al*. 2012). The first reported selective breeding programme for fish was to improve the survival rate to furunculosis in brook trout (Embody and Hayford 1925). Since then many selection experiments have been conducted with a view to improving the growth performance and disease resistance (Kuzema 1971; Kincaid *et al*. 1977; Bondary 1983; Newkirk and Haley 1983; Hetzel *et al*. 2000; Argue *et al*. 2002; Hussain *et al*. 2002;

Langdon *et al*. 2003; Nell and Hand 2003; Gitterle *et al*. 2006; Dunham 2007). Family based selective breeding programme was first started for salmonids in the 1970s (Gjedrem 1985), then for Nile tilapia in 1987 (Ponzoni *et al*. 2010) and for marine shrimp *Penaeus vannamei* in 1993 (Fjalestad *et al*. 1997). Today the number of family based breeding programmes exceeds 100 with the highest number (27) reported for tilapia (Neira 2010; Rye *et al*. 2010).

With the advancement of next generation sequencing (NGS) technology, highly dense marker maps and high-throughput genotyping have become increasingly available in aquaculture species. Genomic selection can increase the genetic gain in aquaculture breeding programs than the traditional breeding programme. Genomic selection is first applied to test population with dense genetic markers followed by predicting breeding values for the selective candidates (Meuwissen *et al*. 2001). In aquaculture, genomic selection might be better to apply within families because the marker densities may not be high in aquaculture species and the number of individuals in full-sib family can be very large (Sonesson and Ødegård 2016).

1.5.3.1 Genetically improved farmed tilapia (GIFT) breeding programme

To revive the tilapia industry and to expand tilapia culture all over the world, it was imperative to produce improved stocks of Nile tilapia, maintaining all the genetic qualities that will produce better performance than their wild predecessors. An effort was made in collaboration with the International Center for Living Aquatic Resources Management (ICLARM, later renamed the WorldFish Center), National Freshwater Fisheries Technology Research Center / Bureau of Fisheries and Aquatic Resources (NFFTRC/BFAR), Freshwater Aquaculture Center / Central Luzon State University (FAC/CLSU), the Institute of Aquaculture Research (AKVAFORSK) and Marine Science Institute of the University of

the Philippines (UPMSI) (Bentsen *et al*. 1998). This was a 10 years long project entitled "Genetic Improvement of Farmed Tilapias" (GIFT) with financial support from the United Nations Development Programme (UNDP) and the Asian Development Bank (ADB), started in April 1988; Philippines as the project site.

To execute the GIFT project, four sampling areas in different African countries (Egypt, Ghana, Kenya and Senegal) were selected. Nile tilapia belonging to two of the different sub-species were collected. Egypt, Ghana and Senegal strains belong to the sub-species *O. niloticus niloticus*, whereas the strain from Kenya belongs to *O. niloticus vulcani* (Trewavas 1983). Four Asian farmed tilapia strains (from Israel, Singapore, Taiwan and Thailand), which were introduced from 1972 to 1984, were also tested along with African strains. The Israel, Singapore and Taiwan strains were thought to originate from Ghana, whereas the Thailand strain came from Egypt (Macaranas *et al*. 1986; Pullin and Capili 1988).

To develop GIFT, an experiment was initiated in 1989 by single pair mating of 7 selected strains (25 breeding pairs from each strain) to estimate the extent of genotype-environment (G×E) interactions for growth, survival and reproductive performance in different environments in the Philippines. Significant differences in growth performances were observed among the strains and most of the African wild strains showed better growth performances compared to the Asian farmed strains. This first generation experimental trial was followed by a complete diallele cross (8×8) experiment involving all strains to produce 64 possible crosses to pinpoint the magnitude of additive and non-additive genetic effects (heterosis). Crossbreeding showed lower growth and survival of offspring over parents (eight strains) and largest non-additive genetic effect was found in only one cross (14 % of mean heterosis). Where the populations are genetically distant, there is a higher possibility of non-additive genetic effects in crosses because of the likelihood of divergent fixation of

interacting alleles (Hill 1982). In addition to this, interactions of non-additive genetic components with environmental variations suggested that exploiting additive genetic component over generations by selective breeding rather than crossbreeding would be an effective way for improving genetic material (Olesen *et al*. 2003). On the basis of additive genetic performance, it was planned to produce a base population of Nile tilapia selecting the best performing pure-bred and cross-bred families from the diallele cross experiments, and following a nested mating design to estimate the genetic gain and response to selection in the third generation. The base population showed considerable genetic variation in terms of growth and survival, which means rapid genetic improvement would be possible. Adopting a combination of "between-family" and "within-family" selection strategy, 20 % more growth was found after one generation of selection than the base population. Selection had been continued for further generations and had been assessed in different test environments (Eknath 1995). Over the first five generations of selection, the genetic gain per generation has been calculated to about 12-17 % and 85 % cumulative genetic gain was found compared to the base population (Eknath and Acosta 1998), which was introduced as the GIFT strain. GIFT was developed with a view to improving growth as a primary trait through selective breeding and has proved to be a landmark development in the history of genetic improvement of tropical fishes. Since the development of GIFT, it has been introduced into different countries to develop breeding programme for special traits, for example hybrid was produced to increase the salinity tolerance and growth rates (de Verdal *et al*. 2014b).

1.6 Sex determination and differentiation in tilapia

Tilapia is getting more attention from fish biologists and geneticists, and has been documented as a model species manifesting a complex sex determination systems

comprising of genotypic, environmental and interaction between genotypic and environment factors (Baroiller and D'Cotta 2001; Bezault *et al*. 2007). Tools such as genetic linkage maps, bacterial artificial chromosome (BAC) based physical map and whole genome sequence for tilapia are now available which will help to search for the master switch(es) to determine their sex with more confidence (Lee *et al*. 2005; Yue 2013).

A wide variety of techniques have been applied to determine the sex of tilapia, and both male and female heterogametic (XX/XY and WZ/ZZ) systems have been found within tilapia species. Interspecies hybridization, chromosome set manipulation and progeny testing following sex-reversal by hormone treatment, reveals that *O. niloticus, O. mossambicus, T. zillii* have a male heterogametic and *O. aureus, O. hornorum, O. karongae, T. mariae* have a female heterogametic sex-determining system (Penman *et al*. 1987; Cnanni *et al*. 2008). Comparison of mitotic karyotypes of tilapia showed that there is no significant difference in chromosome number and chromosome morphology among tilapia species (Majumdar and McAndrew 1986). Construction of linkage maps offers an excellent opportunity for precise identification of the sex-determining region. It was found that genes on linkage group (LG) 1, 3 and 23 determine the sex in tilapiine species (Cnaani *et al*. 2008; Eshel *et al*. 2010, 2012). Sex of blue tilapia (*O. aureus*) is determined by WZ/ZZ system where female (WZ) has two separate pairs of sex chromosomes found by synaptonemal complex analysis. One pair is situated on subterminal region of chromosome 1 while the other is in a small bivalent pair (Campos-Ramos *et al*. 2001). Lee *et al*. (2004) demonstrated that sex of *O. aureus* is determined by the interaction of two unlinked loci. A dominant male repressor (W haplotype) is found in LG3, which is epistatic to a dominant male determiner (Y haplotype) in LG1.

Tilapia hatch at 4 dpf and the yolk sac is absorbed completely at around 9 dpf. They have sexually undifferentiated gonads at this stage and ovary differentiation starts earlier than testis. Gonads are mainly composed of somatic cells and it has 38-58 primordial germ cells (PGC) at 9 dpf. A slight increase in somatic cells and PGCs is found before 20 dpf. Around 20 dpf, higher number of PGCs is found in the future ovaries of XX females. The number of PGCs continues to grow and approximately 217 PGCs in XX females are observed whereas 79 PGCs in XY males around 25-35 dpf. Around 28 dpf ovarian PGCs enter into meiotic prophase and grow progressively to form the future ovarian cavity. On the other hand testicular germ cells enter into meiosis around 55 dpf and initiate spermatogenesis around 90 dpf. The male and female sex differentiation pathway in tilapia is shown **Figure 1.2**.

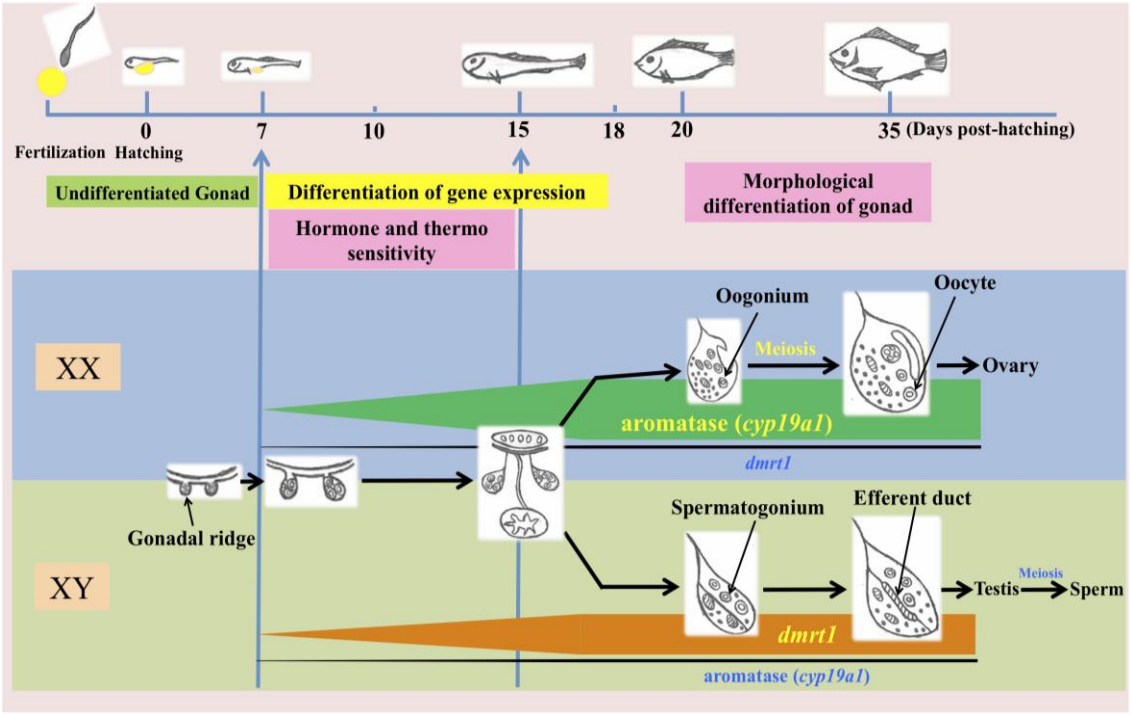


Figure 1.2 Gonadal (male and female) differentiation pathway in tilapia (Original image, Ijiri *et al*. 2008).

1.7 Sex control in fish culture

1.7.1 Importance of controlling sex-ratio in tilapia culture

Sex control in fish species has many potential applications for the aquaculture industry. These include: (i) prevention of precocious maturation and uncontrolled reproduction (e.g. in tilapia), (ii) production of monosex populations due to sex-biased differences in growth rate and commercial value (e.g. tilapia, shrimp), (iii) increasing the stability of mating systems (e.g. sex change in groupers).

Tilapia has a fast growth rate, has a short generation time and is easy to breed in captivity all the year round. In the tropics, tilapia reaches sexual maturity and begins to reproduce at an age of five months while in the subtropics maturity can take a little bit longer depending on the water temperature. Females can breed spontaneously (every 3-4 weeks) during the

breeding season after attaining their sexual maturation. When both sexes are cultured together unwanted reproduction can cause prolific breeding resulting in the overpopulation and stunting of growth in production ponds, which in turn reduces the production and the quality of the products. To overcome this situation, single sex tilapia culture has been preferred in many commercial aquaculture systems. Male tilapia is favoured not only to control the overcrowding problems created by rapid reproduction but also for growth (male grows 20 % faster and has lower feed conversion ratio than female) (Dunham 2011).

1.7.2 Possible ways of controlling sex-ratio

1.7.2.1 Inter-specific hybridization

Interspecific hybridization is the process of combining genetic material from two different species within the same genus. It has been used in a number of fish species primarily as a means of improving production traits (such as growth rate, survival, disease resistance) as well as to manipulate sex-ratio. Rosenstein and Hulata (1994) found that hybridization between Nile tilapia and Blue tilapia produce predominantly male offspring and reduce unwanted reproduction in the culture system. Different sex-determining mechanisms (Nile tilapia has XX/XY and Blue tilapia has ZW/ZZ system) in those two species may induce predominant male production. Hybrids from other crosses like Nile tilapia and Wami tilapia (*O. hornorum*), Mozambique tilapia (*O. mossambicus*) and Wami tilapia are predominantly males. On the other hand 100 % females can be obtained from hybrids between Striped bass (*Morone saxatilis*) and Yellow bass (*M. mississippiensis*) (Wolters and DeMay 1996).

1.7.2.2 Sex-reversal

Skewed sex-ratio in the culture system can be achieved by manipulating the water temperature or administration of sex steroids during the labile period of sex determination. The phenotypic sex of the fish can be changed without changing their genetic composition. High temperature treated fish with ZW chromosomes (genotypically female) may exhibit testis formation as well as other male-specific secondary sexual characteristics (phenotypically male). It is possible to produce phenotypically male fish (genetically female) following androgen treatment (called masculinisation) and phenotypically female fish (genetically male) following estrogen treatment (called feminisation). The first successful artificially induced sex reversed in fish was in the medaka through the administration of estrogens and androgens to sexually undifferentiated fish, and resulted in both functional females and males, respectively (Yamamoto 1953, 1958). He stated that effective sex-reversal varies between species and also the nature of the steroid hormone. He also mentioned that the sex steroids should be given prior to any sign of gonadal differentiation.

Environmental factors can also change the sex-ratio in tilapia species. The critical period for tilapia sex differentiation is from 9 to 15 days post fertilisation (dpf) (D'Cotta *et al*. 2007; Ijiri *et al*. 2008). During this period, temperature or hormonal treatments can alter the sex-ratio. Nile tilapia experience a wide range of habitats with strong seasonal temperature variations from hydrothermal hot springs ($\geq 40^{o}$C) to constant cold temperature (17-24^{o}C) (Admassu and Casselman 2000). Baroiller *et al*. (1995) first described that tilapias are sensitive to change in temperature during the critical period of sex differentiation. It is possible to produce complete sex reversed male (XX) Nile tilapia by rearing at high temperature ($> 34^{o}$C) giving functional male phenotypes (Baroiller *et al*. 1995; Baroiller *et*

al. 2009b; Poonlaphdecha *et al*. 2013). In the case of *O. mossambicus*, larvae exposed at 20^{o}C after 10 dpf induced female skewed sex-ratio. On the other hand exposure of high temperature (32^{o}C) after 10 dpf induced a high proportion of males (Wang and Tsai 2000). Male proportion (97.8 %) of blue tilapia is also increased dramatically when larvae are treated at higher temperatures (34^{o}C) during the sexual differentiation period (Desprez and Melard 1998). Temperature sensitivity correlates with the gonad sensitivity through pituitary gland and gonad axis, like as other abiotic factors such as hormones. Hormonal treatments during gonadal sex differentiation can inverse sex and produce functional phenotypes (Baroiller *et al*. 2009a; Guiguen *et al*. 2010). Genetic and environmental interaction can also alter the sex-ratio in tilapia when exposed to high or low temperature (Baroiller *et al*. 2009b).

1.7.2.3 Chromosome set manipulation

Manipulation of the ploidy levels is one of the most studied approaches to analyse sex determination in aquaculture species and perhaps to change the sex-ratio in experimental studies, but little used in commercial aquaculture. Chromosome set manipulation methods include induced gynogenesis, androgenesis, and polyploidy. This technique restricts the normal developmental processes of gametogenesis and early post-fertilisation and changes the ploidy level through the application of physical or chemical shocks. Depending on when these external shocks are applied, individuals can be haploid (n), triploid (3n), or tetraploid (4n), as compared to the normal diploid (2n) chromosome number.

Gynogenesis is a type of induced all-maternal origin, while androgenesis results in all-paternal origin (apart from the mtDNA). In the case of induced gynogenesis, eggs are fertilised by sperm whose genetic materials are inactivated (mostly by ultraviolet irradiation) before fertilisation and subsequent diploidisation by the suppression of the

second meiotic division (meiotic gynogenesis) or suppression of first mitotic division by physical shocks (either heat, cold or pressure) or chemical treatments. In induced androgenesis, genetically inactivated eggs (mostly by UV or 60Co) are fertilised by genetically active sperms and subsequent diploidisation by suppressing the first cleavage or by fertilising the genetically inactivated eggs with sperms from tetraploid males. Recent efforts have been focused towards the production of triploids and tetraploids to control the maturation using chromosome set manipulation technique (Budd *et al*. 2015). Although there has been a lot of research on producing mono-sex individuals with only a maternally (gynogens) or paternally (androgens) derived set of chromosomes, these techniques have had very little impact on commercial production.

1.7.2.4 Marker-assisted selection

Quantitative trait loci (QTL) are the chromosomal regions (single gene or cluster of gene) that determine a trait of interest. Traits such as growth, flesh quality, sexual maturation, sex determination, disease resistance, salinity tolerance and temperature tolerance are the most important for aquaculture species. These traits can be genetically improved through simple selective breeding (Naish and Hard 2008). QTL for economically important traits have been mapped for more than 20 aquaculture species (Yue 2013), QTL for most of the important traits for tilapia have already been developed (Cnaani *et al*. 2003, 2004; Lee *et al*. 2004, 2005; Rengmark *et al*. 2007), which is one of the most important farmed fishes. Traits that are difficult to measure, exhibit low heritability and expressed late in development, marker-assisted selection (MAS) or genomic selection will be the most important method to exploit these type of trait.

Molecular markers that are directly linked to quantitative trait loci have been regarded as useful for MAS programme. MAS is more efficient than conventional selective breeding

because MAS incorporates both genomic and performance records rather than just performance data. MAS can also reduce time and labour efforts in progeny testing in the long run once the haplotype of the improved trait is defined. The essential conditions to implement MAS are the development of useful resource families to evaluate phenotypes and large number of polymorphic genetic markers that are tightly linked to QTL for trait(s) of interest based on QTL mapping (Ozaki *et al*. 2012). MAS is useful in genetic improvement for traits that are evaluated on sibs of breeding candidates, and sibs are phenotyped and genotyped to estimate the marker-trait associations of the markers that are inherited from the parents (Sonesson 2007). Molecular markers have been applied as an effective tool to identify sex-specific or sex-linked markers, and to control sex-ratio in the culture system (Liu and Cordes 2004).

1.8 Molecular aspects in fish sex determination mechanisms

1.8.1 Molecular markers in fish sex determination

Development of molecular markers has revolutionized the field of aquaculture genetics specially for genetic variability study, parentage analysis, species identification, linkage map construction and characterizing the inheritance pattern of trait (QTL). Identification of suitable sex-linked and sex-specific markers can be an alternative approach to progeny testing to determine the sex of fish. There are two types of molecular marker, protein and DNA based. Protein markers, also known as allozymes determine the allelic variants of protein. Allozyme gel electrophoresis is the co-dominant genetic markers and since 1960s, it has been extensively used in fish molecular genetics (Hillis *et al*. 1996). These markers have been applied for linkage mapping of some salmonids and poeciliids (Morizot *et al*. 1991). Although these markers were studied extensively in population genetics, the low

level of polymorphism and low number of loci were the main barriers to use them in the study of sex determination in fish or other commercial traits.

DNA markers are genes or segments of DNA with a known physical location on a chromosome which is readily detectable and whose pattern of inheritance can easily be monitored (Avise 1994). Although polymorphic (Restriction fragment length polymorphism, RFLP; random amplified polymorphic DNA, RAPD) and highly polymorphic (Amplified fragment length polymorphism, AFLP; microsatellites and single nucleotide polymorphism, SNPs) markers are widely used to identify sex-specific DNA sequences in different fish species, there are still very few reports on sex-linked molecular markers in teleosts. RFLP was the first developed type of co-dominant DNA marker (Botstein *et al*. 1980) based on the digestion of genomic DNA using restriction endonucleases. Linkage maps have been constructed primarily for studying QTL by using RFLP technique (Lander and Botstein 1989). Two male-specific DNA markers have been identified for the first time from African catfish (*Clarias gariepinus*) using random amplified polymorphic DNA (RAPD) technique (Kovács *et al*. 2000).

Highly polymorphic markers (AFLP, microsatellite, SNPs) have been widely used to analyse sex determination in fish because of their abundance and ease of scoring. AFLP, a dominant polymerase chain reaction (PCR) based multi-locus fingerprinting technique was first developed by Vos *et al*. (1995). The main principle of this technique is restriction digestion of genomic DNA followed by ligation of known adapter sequence. Restriction fragments are amplified by PCR and the amplified fragments are separated by gel electrophoresis. AFLP markers have also been studied for the analysis of fish gynogens and androgens (Felip *et al*. 2000) and for the construction of high-resolution linkage map (Blears *et al*. 1998). Sex-linked AFLP markers are species-specific and markers linked to

sex of one species may not be linked to sex in another. Sex-linked AFLP markers have been isolated from Threespine stickleback (Griffiths *et al*. 2000); Medaka (Naruse *et al*. 2000); Nile tilapia (Ezaz *et al*. 2004); Rainbow trout (Felip *et al*. 2005); Half-Smooth tongue sole, *Cynoglossus semilaevis* (Chen *et al*. 2007) and Black tiger shrimp, *Penaeus monodon* (Staelens *et al*. 2008).

Microsatellites are the simple sequence tandem repeats ranging from 1 to 6 base pairs (Litt and Luty 1989). Microsatellites are evenly distributed throughout the genome and occur once in every 10 kb in fishes (Wright 1993). Most microsatellite loci are relatively small, ranging from a few to a few hundred repeats. Ease of scoring, reproducibility, high levels of polymorphism and co-dominant mode of inheritance have made microsatellite marker superior to others. Although microsatellites are extensively used in the construction of linkage maps and to identify the sex-specific DNA sequences in both plant and animal. Linkage maps using microsatellite markers are available for only a few fish species, including Tiger pufferfish (Kai *et al*. 2005), European sea bass (Chistiakov *et al*. 2005), tilapia (Lee *et al*. 2005), Rainbow trout (Sakamoto *et al*. 2000), Salmon (Gilbey *et al*. 2004), Turbot (Bouza *et al*. 2007), Barramundi (Wang *et al*. 2007a), Channel catfish (Kucuktas *et al*. 2009), Grass carp (Xia *et al*. 2010) and Asian sea bass (Wang *et al*. 2011).

SNPs are the single base differences in the DNA between individuals. Like microsatellites, SNPs are distributed throughout the genome, approximately 1 in every 100 bases. SNPs are the most abundant type of genetic marker and their high density makes them ideal for studying the inheritance in genomic regions (Stickney *et al*. 2002). The recent SNP/GBS-based ones are generally much denser (and the microsatellite-based one are often denser than those based on AFLPs, RAPDs etc.). Sex-specific linkage maps of Atlantic salmon have been constructed using SNPs spanning a total of 2402.3 cM in female and 1746.2 cM

in male salmon (Lien *et al*. 2011). Next generation sequencing techniques accurately generate large numbers of SNPs in a very short time making them a powerful tool for genome mapping and identification of quantitative trait loci.

1.8.2 Next generation sequencing (NGS) approaches in fish sex determination

Since 2005, NGS technologies, alternatively called massively-parallel sequencing or multiplex cyclic sequencing, have provided an increasingly cost-effective genome sequencing platform in the new era of high-throughput genomic analysis to accelerate the genomic research on both plants and animals (Mardis 2008a, b). NGS technologies offer novel and rapid ways of genome research in multiple areas like whole genome resequencing from microbes to humans, targeted genome sequencing, genome-wide characterization and profiling of mRNAs, small RNAs, metagenomics, transcriptomics, mapping structure of chromatin and DNA methylation patterns etc. NGS technology can generate hundreds of megabases to gigabases of nucleotide sequence output in a single instrument run which reduce the cost of DNA sequencing dramatically (Voelkerding *et al*. 2009). NGS platforms include Illumina, Ion, Pacific Biosciences (PacBio), Roche 454 and SOLiD. Illumina instruments (HiSeq, Nextseq, MiSeq etc.) are widely used in sequencing applications such as assembly, resequencing, transcriptome, SNP detection and metagenomic studies. Among them, Illumina MiSeq produces significantly fewer reads but its read length (up to 600 bp) is significantly high. On the other hand, PacBio has the ability to produce longer reads with uniform coveage and the average length of the reads is ~10 kb which is significantly longer than any other sequencing platform (Rhoads and Au 2015).

RFLP, RAPD, AFLP (previously used), microsatellite (less likely to be used now) markers were used to construct linkage maps but they are expensive, sometimes problematic and

difficult to get sequence based marker informations (Baxter *et al*. 2011). Next generation sequencing techniques offer rapid generation of linkage maps consisting of thousands of co-dominant sequenced markers linked to the genes of interest across the genome in a single experiment. Whole genome sequencing will also help to identify genetic variants like single nucleotide variants (SNVs) or SNPs, small insertions and deletions (indels, 1–1000 bp), and structural and genomic variants (> 1000 bp) (Davey *et al*. 2011; Miller *et al*. 2012).

1.8.3 Restriction-site associated DNA sequencing (RADseq) technique

The restriction-site associated DNA (RAD) sequencing was first proposed by Miller *et al*. (2007), a reliable reduced representation of the genome sequencing method, with a view to reducing genome complexity analysis. This method relies on restriction enzyme digestion of genomic DNA following an adapter ligation containing a molecular identifying sequence (MID) unique to each sample and sequencing the DNA associated with each restriction site using the massively parallel Illumina sequencing technology (Baird *et al*. 2008; Davey and Blaxter 2010). This method is being widely used in genetic marker discovery, genome assembly, linkage mapping and QTL mapping. RADseq was first applied for genetic marker discovery in Threespine stickleback whose genome has already been sequenced (Hohenlohe *et al*. 2010). Two types of markers can be generated from RADseq i.e. SNPs and simple sequence repeats (SSR/microsatellite) (Pfender *et al*. 2011). RADseq is less expensive than whole genome sequencing and it has some other attractive features; 1) it sequences nucleotides next to restriction sites and detect SNPs; 2) selection of suitable restriction enzyme helps to get a good number of markers and the number of markers can be increased almost indefinitely by using additional enzymes and 3) it is possible to apply bulk segregant analysis (BSA) to RADseq for genotyping pools of individuals. RADseq technique has been applied to construct genome maps and to identify traits of interest in different fish species.

This technique has been used to study sex determination in different fish species like Atlantic salmon (Houston *et al*. 2012), Spotted Gar (Amores *et al*. 2011), Sockeye salmon (Everett *et al*. 2012), Nile tilapia (Palaiokostas *et al*. 2013a), Atlantic halibut (Palaiokostas *et al*. 2013b). The steps of constructing the RAD library and sequencing are given in **Figure 1.3**.

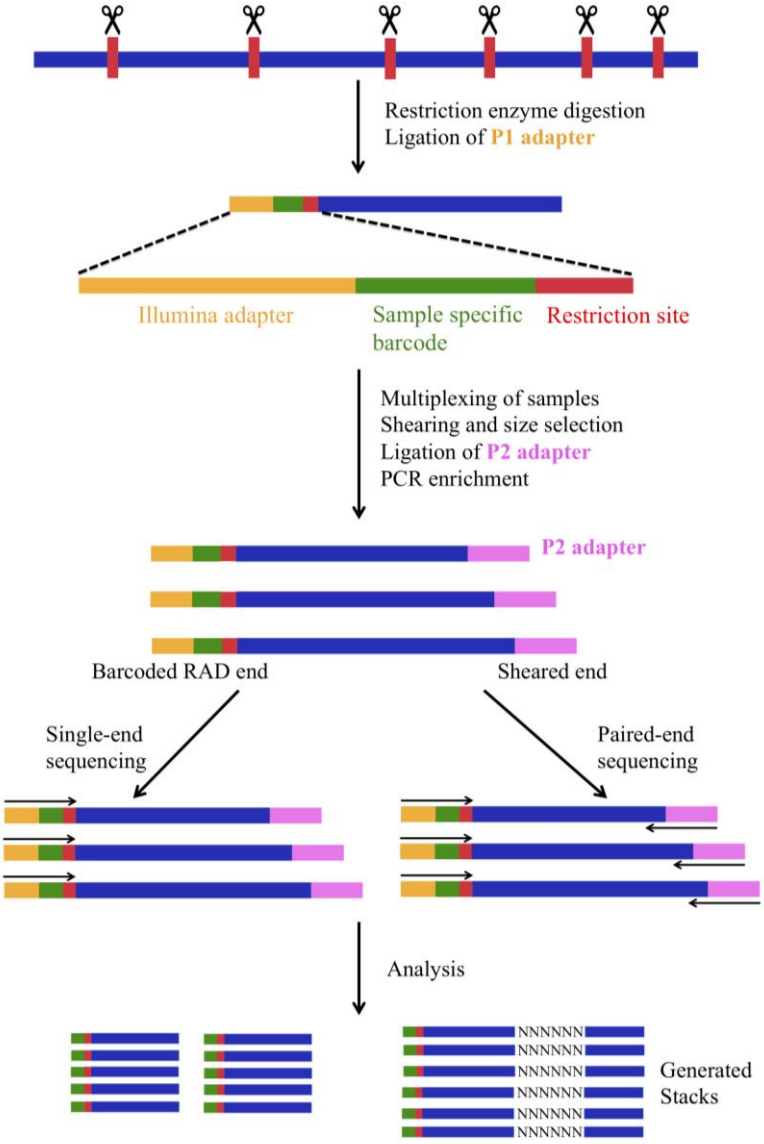


Figure 1.3 The principle behind the standard restriction-site associated DNA (RAD) sequencing (Original image, Baird *et al*. 2008).

With the modernization of molecular genetic research, scientists have identified some limitations for novel RADseq such as it uses one restriction enzyme, random shearing to produce fragments and it has little control over the fragments that are sequenced. For organisms without a reference genome, a significant portion of the RADseq data has been discarded due to sequence read errors and the presence of variable sites (Hohenlohe *et al*. 2011; Pfender *et al*. 2011). Double digest RADseq (ddRADseq), a variation of RADseq, was designed to overcome the RADseq weaknesses. It eliminates random shearing and uses two restriction enzymes to digest genomic DNA followed by adapter ligation and finally specific size selection before sequencing. This method is inexpensive, rapid, requires little starting material (i.e. 20 ng of DNA), does not require prior genomic information and is suitable for high-throughput applications (Peterson *et al*. 2012). ddRADseq has been used to identify the sex-determining region in fish species such as Nile tilapia (Palaiokostas *et al*. 2015); Hāpuku, *Polyprion oxygeneios* (Brown *et al*. 2016).

1.8.4 Bulk segregant analysis (BSA) with different molecular techniques

BSA is a rapid QTL mapping technique for identifying genomic regions containing genetic loci affecting a trait of interest. BSA was first developed by Arnheim *et al.* (1985) to detect the loci associated with human disease and subsequently BSA approach was used to study disease resistance in plants (Michelmore *et al*. 1991). The BSA concept was the phenotypic changes must be reflected by changes in their genotypes. It involves screening for phenotypic differences between two pooled DNA samples derived from a segregating population that originated from a single cross (Michelmore *et al*. 1991). Sometimes detection of individual variation is difficult and in that situation pooled samples (bulk) can be another option to detect a specific gene of interest. Pools can be prepared in two different

ways such as combining equal concentration of DNA for each individual or combining equal amounts of tissue/cells from individuals followed by DNA extraction.

BSA has been combined with various types of molecular markers including RFLP (Palti *et al*. 1999), RAPD (Iturra *et al*. 1997), AFLP (Lee *et al*. 2011), SSR/microsatellite (Lee *et al*. 2003; Wang *et al*. 2014). BSA with different molecular techniques has been used to determine the sex of different fish species (Iturra *et al*. 1997; Lee *et al*. 2003, 2004; Ezaz *et al*. 2004; Keyvanshokooh *et al*. 2007; Wang *et al*. 2009a, b; Lee *et al*. 2011).

With the advent of NGS technologies, BSA has been combined with different NGS platforms to cost-effectively studying lots of individuals from a single population or from different populations in a single sequencing run compared to the sequencing of individual samples. Pooled samples provide more accurate allele frequency estimation at a lower cost than sequencing of individuals (Futschik and Schlötterer 2010). BSA-NGS is more suitable when the experiments rely on the analysis of large number of samples but it has lots of other applied sides such as genotype-phenotype mapping, molecular ecology, genome evolution, cancer genomics etc. BSA has now been combined with the sequence-based analysis such as RADseq, exome sequencing, RNAseq and whole genome sequencing (Schlötterer *et al*. 2014). The concept of BSA-RADseq is shown in **Figure 1.4**. A combination of BSA and NGS techniques are now widely used in plants and microbial genetics. Very few studies have been done in fish so far such as Threespine stickleback (Baird *et al*. 2008) and Channel catfish (Wang *et al*. 2013). Although BSA-NGS has lots of attractive features, it is not sensible to study endangered species because the idea behind BSA is to use lots of individuals in a single pool. Different representation of individuals (DNA concentration) in the pool sometimes interferes in the BSA-NGS results.

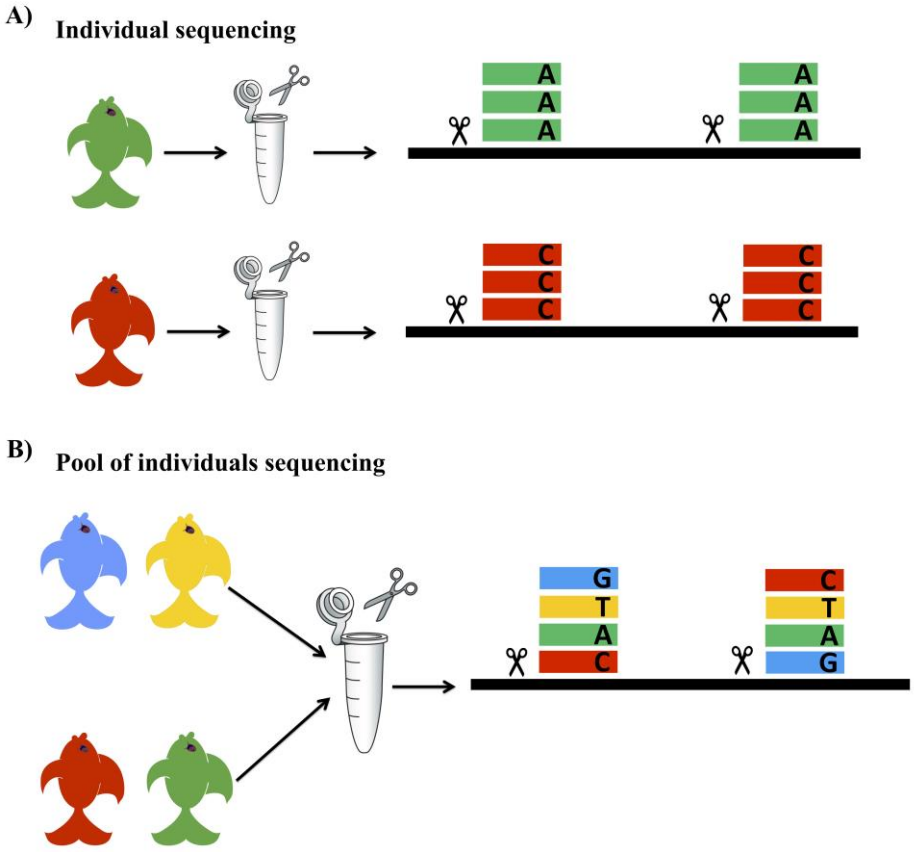


Figure 1.4 Sequencing of pools of individuals compared to the individual sequencing using RAD sequencing (Original image, Schlötterer *et al*. 2014).

1.9 Genome mapping of tilapia to determine sex

Genetic mapping allows accurate location of genes that can be used in programs of genetic improvement in aquaculture, often through linked genetic markers. Genetic linkage mapping is the linear ordering of markers along the chromosome. Linkage mapping is an essential tool for both plants and animals whose genome is yet to be sequenced. Genetic linkage maps are constructed by mapping polymorphic DNA markers based on their segregation relationship. Once the linkage map has been constructed, it serves as a starting point to identify markers that are closely associated with quantitative trait loci (QTL) for target traits.

Sex in tilapia is primarily determined by major sex-determining loci, with other loci with small effects and environmental factors also being involved. Several sex-linked markers have been identified in *O. niloticus* (Nile tilapia) and *O. aureus* (Blue tilapia), and mapped to different linkage groups (LGs) (Lee *et al*. 2003, 2004; Shirak *et al*. 2006; Cnaani *et al*. 2008; Eshel *et al*. 2010, 2012). Three sex-determining loci have been mapped on LG1, LG3 and LG23 in tilapiine species. In purebred *O. niloticus* (XX/XY system), QTL for sex determination were detected on LG1 and LG23 (Lee *et al*. 2003; Eshel *et al*. 2010, 2012). Whereas, in hybrids (*O. niloticus* x *O. aureus*), a QTL for sex was mapped on LG3 (Lee *et al*. 2005). In Blue tilapias (WZ sex-determining system) sex has been found to be controlled by a major QTL on LG3. Lee *et al*. (2004) found that three microsatellite markers *UNH168*, *GM271* and *UNH131* in LG3 were located within a few centimorgans of the sex-determining locus. For Red tilapias, a major sex-determining locus was found in LG22 (Liu *et al*. 2013). A very recent study has identified a gene called *OsZfand3* (Zinc finger AN1-type domain 3) in LG1 in hybrid tilapia (crosses from Mozambique and Red tilapia), which is thought to be tightly linked to the sex-determining locus (Ma *et al*. 2016).

A first generation linkage map of Nile tilapia was constructed using 62 microsatellite and 112 AFLP markers, and consisted of 30 linkage groups spanning 704 cM (Kocher *et al*. 1998). Lee *et al*. (2005) developed a second generation linkage map of Nile and Blue tilapia by using 525 microsatellite markers and found 24 linkage groups spanning 1311 cM. Another linkage map of Mozambique and Red tilapia was constructed using microsatellite markers and consisted of 22 linkage groups spanning 1067.6 cM (Liu *et al*. 2013). In Nile tilapia two microsatellite markers (*UNH995* and *UNH104*) on LG1 have been found to be closely located to a major sex-determining locus (Lee *et al*. 2003). On the other hand Eshel *et al*. (2011) described a strong association of the microsatellite marker *UNH898* (LG23) with the phenotypic sex in mixed sex (XX/XY) populations of *O. niloticus* and *UNH898*

has been mapped next to *Amh* - 1 cM distance (Shirak *et al*. 2006). *Amh* showed strong sexually dimorphic expression in brains of *O. niloticus* at 14 dpf shortly before the initial start of the gonadal differentiation (Poonlaphdecha *et al*. 2011). Eshel *et al*. (2012) also reported that a major sex-determining locus for Nile tilapia is located between microsatellite markers *ARO172* and *ARO177* in LG23. Family specific sex QTLs have been identified in LG1, 3 and 23 which are related to temperature dependant sex-reversal in Nile tilapia (Lühmann *et al*. 2012).

A very good quality Nile tilapia genome has been sequenced from female fish and is publicly available, which will help to find reliable markers related to QTL (Brawand *et al*. 2014). Palaiokostas *et al*. (2013a) constructed a linkage map of Nile tilapia using RADseq with 3,280 informative SNPs. They identified a major sex-determining region in LG1 and found 2 SNP markers (*Oni23063* and *Oni28137*) that are closely associated with the phenotypic sex. Palaiokostas *et al*. (2015) also identified a new sex QTL in LG20 which causes sex-reversal (masculinisation) in Nile tilapia with respect to the XX genotype at the major sex-determining locus in LG1, particularly when fry are reared at elevated temparature. A novel male-specific duplication of *Amh* has been identified by Eshel *et al*. (2014). Male-specific duplication of this gene, denoted by *Amhy*, differing from the sequence of X-linked *Amh* by a 233 bp deletion on exon VII and the lack of transforming growth factor beta receptor (TGF-β domain). This gene has been mapped to the QTL region in LG23 and has found to have potential role in Nile tilapia sex determination. An allelic variant has been identified in *Amh* - exon VI, which was found to have sex-determining role in temperature sensitive families of Nile tilapia (Wessels *et al*. 2014). Later on Li *et al*. (2015) isolated a Y-specific two duplicated copies of *Amh* gene, designated as *AmhΔy* (Eshel *et al*. 2014 called as *Amhy*) and *Amhy*. The *Amhy* is a tandem duplicate copy located immediately downstream of *AmhΔy* on the Y chromosome and the coding sequence of

Amhy is identical to the X-linked *Amh* except a missense SNP (C/T) and 5608 bp lacking of promoter sequence. That SNP was thought to have a critical role in male sex determination in Nile tilapia.

From the overall discussion, it can be postulated that at least two major sex-determining loci located in two different linkage groups (LG1 and LG23) control the sex in Nile tilapia but their interaction or the actual responsible gene(s) remains unknown.

1.10 Aims of the present study

The use of single sex production systems and proper genetic management of the tilapia can significantly increase the global production. Nile tilapia has been the subject of much research interest, inlcuding the complex sex-determining system(s), loss of purity due to introgression, and poor genetic management. Given the evidence of different sex-determining loci in different populations of Nile tilapia, this research was designed to study the sex determination in Nile tilapia at the population level, which will help to apply more efficient MAS to control sex-ratio in production systems. Besides this, the research also focused on the use of genomic tools to assess and maintain the genetic (species) purity of tilapia stocks. In brief, the PhD research was based on the following objectives:

1. Development of a minimally invasive DNA sampling method from fish mucus for standard genotyping techniques (microsatellite and SNP).
2. Verifying the suitability of DNA from fish skin mucus for next generation sequencing (ddRADseq).

3. Methodology development to allow screening of large number of individuals in a single population or multiple populations by combining bulk segregant analysis with double-digest restriction-site associated DNA sequencing (BSA-ddRADseq).
4. Analysing sex determination in genetically improved farmed tilapia (GIFT) using BSA-ddRADseq and validation with different DNA-based markers.
5. Given the emerging evidence of variation in sex determination between and within populations of Nile tilapia, analysis of the Stirling Nile tilapia populations with LG1, LG20 and LG23 markers linked to known sex-determining loci will be done.
6. Investigation of hormonal feminisation in Nile tilapia using combined estrogen hormone and high temperature treatment.
7. Identification of different species and strains of tilapia using species-specific SNP markers.

Chapter 2

General materials and methods

The Stirling population of Nile tilapia and the genetically improved farmed tilapia (GIFT), a strain of Nile tilapia developed through selection from a synthetic base population, were used in this study. Fin tissues of the GIFT strain were supplied by the WorldFish Center, Malaysia. Molobicus hybrid (developed by crossing GIFT and feral Mozambique tilapia followed by selective breeding), hybrid generation 07 and pure species of Nile, Mozambique and Blue tilapia were also used in this study. Experiments on Nile tilapia were conducted at the Institute of the Aquaculture, University of Stirling, UK. Therefore handling of live fish was only required for Stirling Nile tilapia.

2.1 Basic maintenance of Nile tilapia (*Oreochromis niloticus*) and handling procedure

An accredited training under the Animals (Scientific Procedures) Act 1986 (ASPA) was followed and a Personal Licence (I6BF2C636) was obtained before any experimental work was carried out with fish. The basic maintenance of all the experimental stocks followed working procedures under ASPA throughout the study and was monitored periodically by the Home Office. Experimental work was carried out under a Project Licence held by David J. Penman.

All the experimental fish were maintained in recirculating aquaculture systems (RAS) in the Tropical Aquarium Facilities (TAF) of the Institute of Aquaculture, University of Stirling. These systems had a controlled temperature environment with proper facilities for filtering and purifying the water before recycling back to the fish. The water used to flush out the fish waste and water lost due to evaporation was replaced with new water to the systems (heated and aerated water in the case of regular cleaning of the filter tanks, plus low volume continual top up with tap water). The water quality parameters, particularly dissolved

oxygen, ammonia, nitrate and nitrite contents were monitored on a weekly basis. The standard water temperature in the TAF was maintained at around 28^{o}C. Handling of fish was done with proper care. Fish were captured using nets of proper mesh sizes and were then transferred to plastic buckets with water (approximately 28^{o}C) from the aquarium. Nets were disinfected using iodophore (major components phosphoric acid, iodine and non-ionic surfactant). Fishes were anaesthetised with the recommended dose of anaesthetising agents. For anaesthetisation, a stock solution was first prepared by dissolving benzocaine powder (ethyl-4-aminobenzoate, Cat. No. E1501 Sigma Aldrich, UK) at 10 % (w/v) in ethanol and the working solution at a final concentration of 1:10,000 was used to anaesthetise fish.

For keeping the stock record, broodfish were tagged using a 10 digit TROVAN Passive Integrated Transponder (PIT) tag. Before tagging, the needle (attached to a syringe) and the PIT tags were disinfected in 70 % ethanol for 5 min. Anaesthetised fish was kept on wet tissues and the needle with PIT tag was inserted through the body wall on the lateral abdominal side. As soon as the needle penetrated the body wall, the tag was injected carefully to avoid the penetration of air and the syringe was removed with proper caution. The fish was immediately transferred to clean aerated water for recovery and was moved back to the tank.

Broodfish were held in square fibreglass tanks (generally male) or in glass aquaria (generally female). In the glass aquarium, it is easy to see the swollen reddish urogenital papilla of the female, which is the sign of readiness to spawn (ovulation). When the female was ready to spawn, following anaesthetisation ovulated eggs were stripped. Gentle pressure on the ventral abdomen allowed ovulated eggs to come out into a Petri dish. Immediately after stripping, the spent fish was kept in a bucket with aerated water until recovery. Stripped eggs were washed several times to remove any faeces, mucus or scales. Milt was stripped from an anaesthetised male into the Petri dish containing the eggs, by

gentle pressure on the abdomen. Following stripping of milt, the male was held in aerated water until recovery. Eggs and milt were stirred and left for a few minutes to ensuring maximum fertilisation rate. The stripped fish were returned back to their original tanks after recovery.

Fertilised eggs were washed several times and transferred to a plastic down-welling incubator in a recirculating system with UV irradiation of water immediately before the egg incubators. The water flow in the incubator was maintained via a narrow plastic tube in such a way that the fertilised eggs in the incubator were kept in continuous gentle motion, mimicking maternal oral incubation. Fertilised eggs were kept in the incubator for about 10 days upto the point of yolk sac absorption and were ready to transfer for first feeding. The number of fry was counted before transfer and any dead individuals were removed.

Fry were then reared at a density of 50-80 per tank for 3-4 months, either in a 20 L tank in a recirculating system, or if elevated temperature and/or steroid hormone treatments were to be applied (see section 2.2) in a static 5 L tank for the treatment duration followed by subsequent rearing in a recirculating system.

2.1.1 Feeding techniques

Commercially available trout feed (Trout Aquaculture Nutrition, UK; manufacturer Skretting, Preston, UK; Product code 470405) with different sizes (designated as no. 3-5) were used to rear fish in the TAF. Feed ingredients were wheat, soybean meal, maize gluten (60 %), fishmeal, sunflower meal, fish oil, minerals, vitamins.

Proximate nutrient composition of the supplied feeds was as follows:

Trout feed 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).

Trout feed no. 4: 18 % oil, 9 % ash, 1 % phosphorus, 54 % protein, 1 % fibre, 12000 IU/kg Vit.A, 2000 IU/kg Vit.D3, 250 IU/kg Vit.E as alpha tocopherol acetate, antioxidants (BHT, BHA).

Trout feed no. 5: 8 % oil, 8 % ash, 1.2 % phosphorus, 40 % protein, 2 % fibre, 12000 IU/kg Vit.A, 2000 IU/kg Vit.D3, 100 IU/kg Vit.E as alpha tocopherol acetate, antioxidants (BHT, BHA).

Following yolk sac absorption (10 days post fertilisation), hatched fry were fed with powdered feed, prepared by passing feed size 5 (4 mm diameter) through a blender, and fed *ad libitum* for the first four to six weeks. This was followed by feeding a mixture of ground feed and feed size no. 3 as the fish grew. When the fry and fingerlings reached 5 g to 40 g, only feed size no. 3 was used 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.

2.2 Treatment of fry with hormone and temperature

2.2.1 Hormone feed preparation

Feminizing steroid hormones were used for progeny from XX × XY and XX × YY crosses to change the phenotypic sex of XY fish from male to neo-female (genotypically male but phenotypically female). Two different types of steroid hormones were used i.e. Diethylstilbestrol (DES) hormone (Cat. No. D4628, Sigma Aldrich, UK) and 17α-

Ethynylestradiol (EE2) hormone (Cat. No. E4876, Sigma Aldrich, UK). Different doses of hormones were applied based on the objectives of the experiments (**Table 2.1**).

Table 2.1 Amount of hormone required to prepare 100 g of hormone feed based on the doses of hormone.

Hormone	**Dose (mg/kg feed)**	**Amount of hormone (mg) for 100 g of feed**
Diethylstilbestrol (DES)	1000	100
17α-ethynylestradiol (EE2)	100	10
	150	15

Hormones were always protected from light. The required amount of hormone was weighed in a plastic graduated centrifuge tube using an electric balance to prepare 100 g of hormone feed. 30 mL absolute alcohol was added to the hormone and mixed until the hormone was completely dissolved. Feed was prepared by grinding feed size no. 5 and sieved to remove any large particles to make it suitable for first-feeding fry. Pre-weighed 100g of feed was placed in a plastic tray and the hormone solution was added gradually to the feed using a 5 mL micropipette, stirring with a spatula to ensure even mixing. The treated feed was allowed to dry in a fume hood overnight before being packed into a properly labelled plastic bottle and kept in the freezer (-20^{o}C) to avoid degradation of hormone until use.

2.2.2 Experimental set up for the treatment

Two egg troughs (215 cm × 55 cm × 16 cm) were used, one for the normal water temperature (28^{o}C) and the other for the high temperature (36^{o}C) treatment. Each egg trough can accommodate maximum 10 aquaria (5 L capacity for each). The water

temperature was maintained at 37^{o}C by using a sensitive thermostat heater (GD120 Grant, England), with pumps to circulate the water in the egg trough to equalise the temperature for all of the aquaria sitting in it (**Figure 2.1**). The water temperature inside the aquarium increased gradually from 28^{o}C to 36^{o}C in about 2 hrs. Temperature was checked three times a day, given that 36^{o}C is fairly close to the upper lethal temperature. The water temperature for high temperature treatment was 36.11^{o}C ± 0.12 (mean ± SD) throughout the experiments. Waste materials/unused feeds were removed from the aquarium by siphoning twice a day i.e. in the morning and in the afternoon. The water siphoned out (c. 1.0-2.0 L) was replaced. One extra aquarium was kept in each trough as water reservoir to use to replace water in treatment aquaria (maintain the same temperature throughout the experiment). Each aquarium had its own aerator to maintain the proper oxygen supply for the fry.

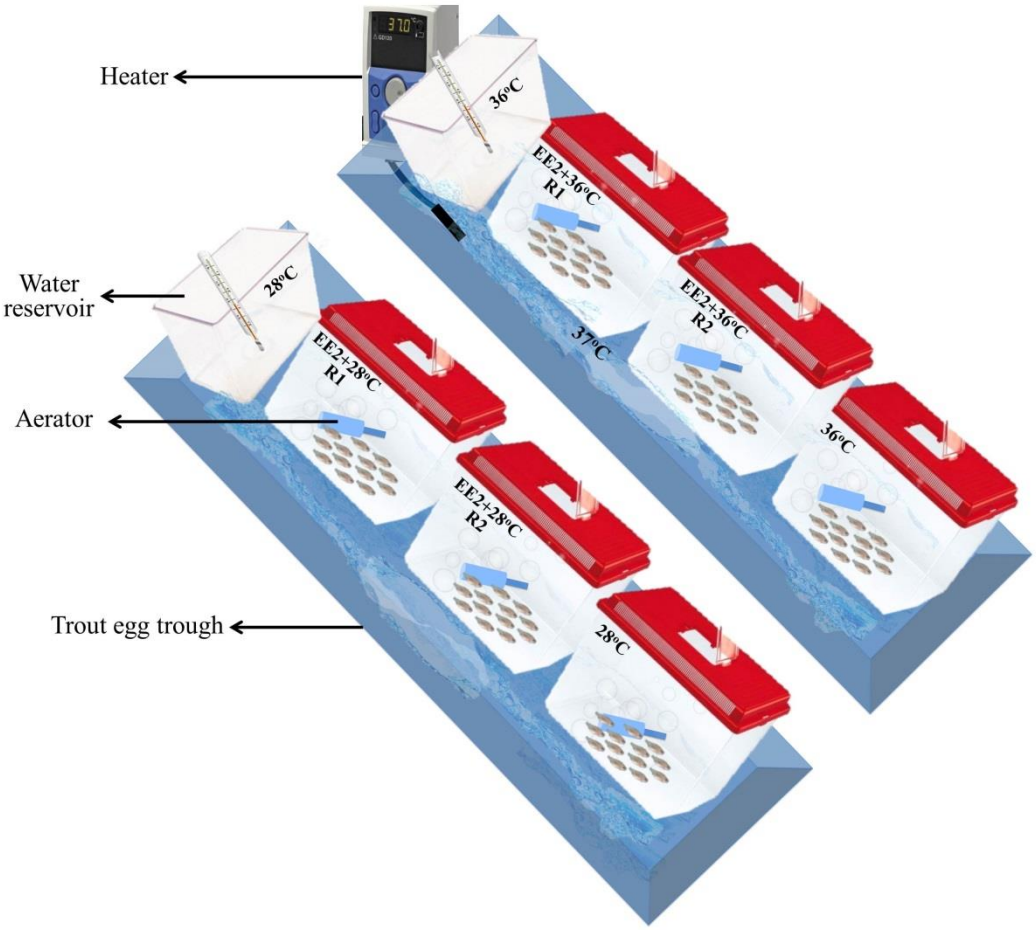


Figure 2.1 Experimental set up for sex-reversal experiment (Original image).

Equal numbers of 10 dpf fry from a single family were split into different groups based on the objectives of the experiments. A control group was also used for each family.

2.2.3 Hormone treatment

For hormone treatment, 10 dpf fry were transferred from the incubator to 5 L plastic aquarium containing water at the same temperature as the incubator (28^{o}C). The aquarium was then placed into a trout egg trough (see description in 2.2.2). Fry were fed with hormone-treated feed five times a day *ad libitum*. After 20 days of hormone treatment, fry

were counted to determine the survival rate and then transferred to a 20 L tank in a recirculating system.

2.2.4 Temperature treatment

Counted 10 dpf fry were transferred in a 5 L plastic aquarium where the water temperature was 28^{o}C. Then the aquarium was placed in a trout egg trough where the water temperature was maintained at 37^{o}C (see description in 2.2.2). Larvae were fed with normal feed *ad libitum* five times a day. After 10 days of high temperature treatment, the aquarium containing the fish was transferred to the next trout egg trough where the water temperature was 28^{o}C, and the temperature inside the aquarium decreased gradually to 28^{o}C. Following counting, fry were transferred to a 20 L tank in a recirculating system.

2.2.5 Combined hormone and temperature treatment

For combined treatment (hormone and temperature), fry were transferred to a plastic aquarium (temperature 28^{o}C), which was placed in the trough where the temperature was 37^{o}C. Larvae were fed with hormone-treated feed five times a day *ad libitum*. After 10 days of combined treatment, either the temperature of the heater was reduced to 28^{o}C or the aquarium was placed into trough where the water temperature was 28^{o}C. The fry were fed with hormone-treated feed for another 10 days and after that the fry were shifted to a tank in a recirculating system following counting.

Following hormone and temperature treatment, fry were reared in a recirculating system for at least three months for phenotypic sex identification. Some of the fish from the sex-reversal experiments were kept alive as potential broodstock.

2.3 Killing and gonad squashing

Phenotypic sex was determined microscopically following squash preparation according to the protocol described in Guerrero III and Shelton (1974). After three months of rearing, fish were killed by an approved Schedule 1 humane killing method: immersion in an overdose of anaesthesia (0.05 % benzocaine) for at least 10 min, followed by destruction of the brain. Whole gonad was excised with a fine forceps and mounted on a microscope slide. A few drops of acetocarmine dye were added and the gonad was squashed with a coverslip. The squashed gonad was examined under a microscope for phenotypic sex identification (**Figure 2.2**). The male gonad was composed of fine granular like spermatogonia and the female was characterized with the structure of circular oogonia. A few intersex fish were also identified, composed of spermatogonia with very few circular oogonia. Following gonad squash, fin samples were also collected and fixed into 100 % ethanol for subsequent DNA extraction and genotypic sex identification.

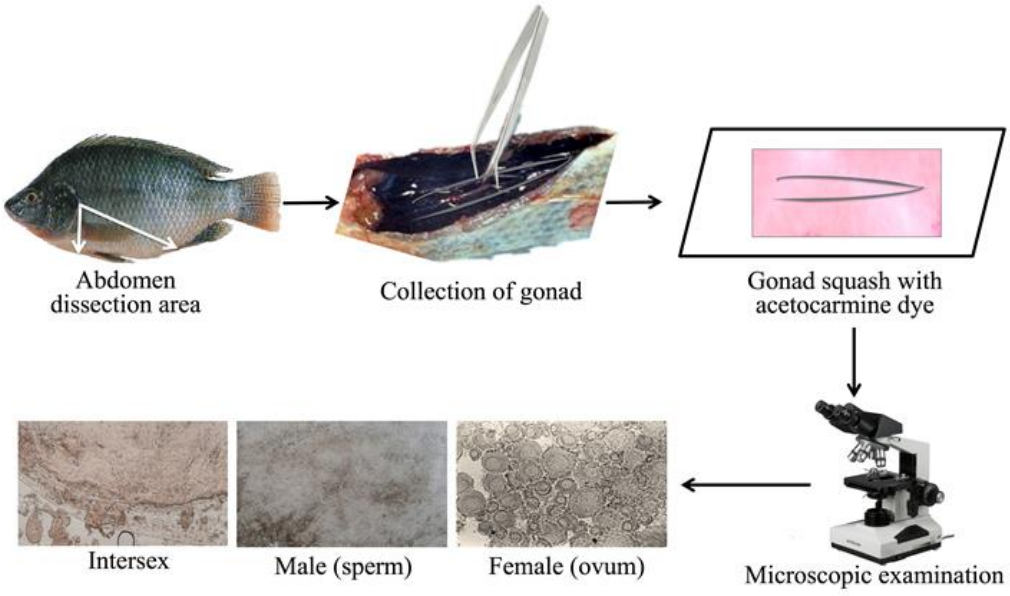


Figure 2.2 Microscopic examination of the phenotypic sex in Nile tilapia through acetocarmine gonad squash method.

2.4 DNA tissue sampling

2.4.1 Collection of fin/muscle tissue

Before fin tissue collection, fish were anaesthetised (with recovery, to allow sampling) using benzocaine in a recommended dose for brood and small fishes. Tissue from caudal or dorsal fin was collected and preserved in fixative (100 % ethanol) in a 2.0 mL screw cap tube. Fish muscle sample was also collected following Schedule 1 humane killing and preserved in 100 % ethanol for mucus/fin/muscle comparison experiment (see chapter 3).

A laser printed sample information label was stored in tube with sample; all information was also recorded in a spreadsheet to aid logistics.

2.4.2 Minimally-invasive DNA sample collection

The minimally-invasive DNA sampling method developed (see Chapter 3) did not require anaesthetising the fish (if anaesthetic had been used, this would have counted as a procedure under ASPA). The fish were held in a bucket with aeration. Each individual fish was netted out. Mucus samples were collected using brush and / or filter paper. A non-sterile brush with a handle (Endocervical sampler, CellPath Ltd, UK) was gently drawn along the flank or rubbed inside the mouth of the fish 10 times to collect skin mucus or buccal mucus respectively. Filter paper (2 cm × 0.5 cm) was also used to absorb skin mucus from the body surface. Immediately after sample collection, the mucus sample on the brush head or filter paper was either preserved in 100 % ethanol and stored in the cold room until use or placed directly into lysis solution for immediate DNA extraction (**Figure 2.3**). Immediately after sample collection, the fish were returned to the original tank.

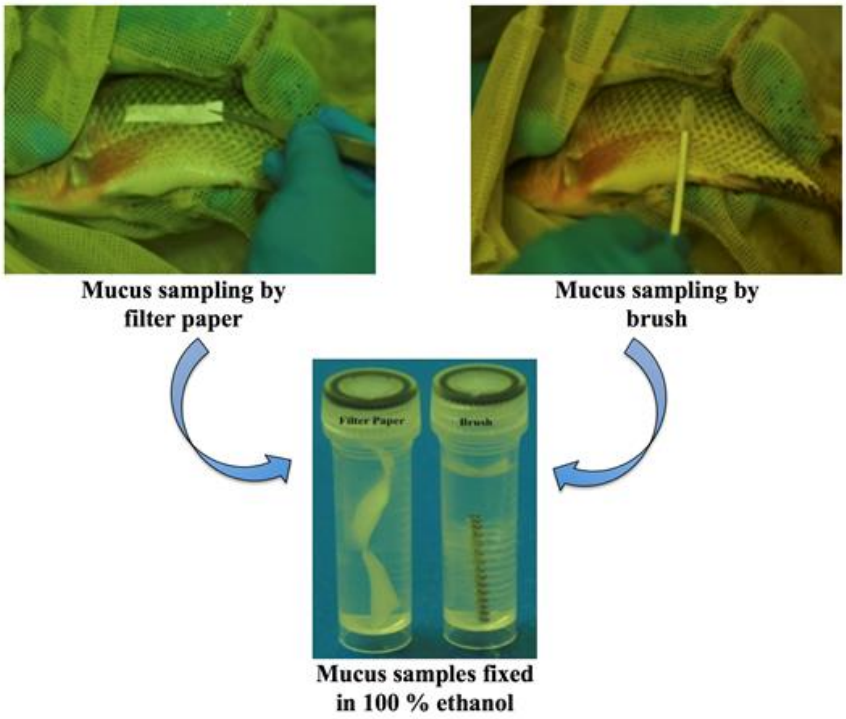


Figure 2.3 Mucus sampling from live fish (Nile tilapia) by filter paper or swab brush, and fixation in ethanol.

2.5 DNA extraction protocols

DNA can be extracted using a range of protocols. The following two protocols were used to extract DNA from collected samples in this research:

i) Protein salting out and isopropanol precipitation (SSTNE/SDS) method

ii) HotSHOT method

The first method was used for fin, mucus and muscle samples with some modifications from Aljanabi & Martinez (1997). This method provides excellent quantity and quality of genomic DNA, which is adequate for standard genotyping, restriction digestion, next

generation sequencing etc. The major advantages of this method are low cost, simplicity and remove the need to use hazardous chemicals (phenol, chloroform etc.).

The HotSHOT method (Truett *et al.* 2000) was used to extract DNA from mucus and fin samples to compare the efficiency of DNA for genotyping. This is one of the quickest DNA extraction methods. Within 30 min DNA is ready for PCR, and large numbers of samples can be extracted in parallel. The quality of DNA is not as good as that from the salting out method, because it still contains protein, RNA and other inhibitors, but the DNA has been shown to be adequate for PCR and standard genotyping. A small amount of tissue is required for extracting PCR perfect DNA. Sometimes we need to increase the amount of DNA (0.5-5.0 µL) for PCR.

2.5.1 Protein salting out and isopropanol precipitation (SSTNE/SDS) method

A stock solution of SSTNE extraction buffer (1 L) was prepared by dissolving the chemicals mentioned in **Table 2.2** using magnetic stirrer, the solution could then be stored in a fridge. Lysis solution (50 mL working solution) was prepared by adding 20 % SDS to SSTNE (2.5 mL SDS and 47.5 mL SSTNE) to get a final concentration of 1 % SDS. Aliquots of 220 µL of lysis solution and 5 µL of 10 mg/mL proteinase K (Sigma Aldrich, UK) were added to a nucleic acid-free 1.5 mL labelled Eppendorf tube for each sample. For fin or muscle samples that had been fixed in ethanol, the sample was placed briefly on tissue paper to remove any excess ethanol. Approximately 20 mg of fin tissue was taken and cut into small pieces and added to the lysis solution. The tubes were incubated at 55^{o}C overnight on a rotator to aid mixing and digestion.

The digested samples were incubated for 15 min at 70^{o}C to inactivate the proteinase K, cooled to 37^{o}C and 5 µL (2 mg/mL) of DNase free RNaseA (Sigma Aldrich, UK) was

added to each tube followed by vortexing. After brief centrifugation, samples were incubated at 37^{o}C for 60 min. Samples were cooled down to room temperature, 160 µL (0.7 × vol.) of 5 M NaCl was added to each sample which was then mixed by vortexing and left on ice for 10 min, which helps to precipitate protein by binding with NaCl. The samples were then centrifuged at high speed (21,000 × G) for 10 min to spin down the precipitated protein. Following centrifugation, the supernatant (250 µL was adequate to get the high quality DNA pellet) was pipetted out without touching the protein pellet and transferred into a clean, labelled Eppendorf tube. An equal volume of absolute isopropanol (250 µL) was added to the supernatant and mixed by 5-6 sharp (rapid and abrupt) inversions or sharp vortexing to precipitate DNA. At this stage, the thread-like DNA structure was normally visible and very brief (5 seconds) centrifugation was then used to pellet the DNA (longer centrifugation time is likely to add impurities to the DNA pellet), if no DNA precipitate was visible, samples were left on ice for 10 min followed by centrifugation for 10 min at 21,000 × G. The bulk of the supernatant was carefully poured off, the tube was briefly centrifuged and as much isopropanol as possible was pipetted out without disturbing the DNA pellet. The DNA pellet was washed twice with 70 % ethanol. 1 mL of 70 % ethanol was added for the first wash and the sample left on a paddle mixer for two hours. After brief centrifugation, the ethanol was poured off, another aliquot of 1 mL of 70 % ethanol was added and the sample left on a paddle mixer overnight.

After the overnight washing, tubes were taken out from the paddle mixer and centrifuged at high speed (21,000 × G) for 5 min. The supernatant was carefully poured off, briefly centrifuged and as much ethanol as possible was pipetted out without touching the DNA pellet. The DNA pellet size and appearance were recorded. The DNA pellet was allowed to air dry for about an hour and depending on the size of the pellet, 15-30 µL of 5 mM Tris (pH 8.0) was added and the sample left at room temperature for an hour to aid dissolving

the DNA pellet. The DNA sample was stored at 4°C for at least one day to allow the sample to dissolve before quantification.

In case of the mucus samples collected by filter paper, the paper was cut into small pieces and added into 300 µL lysis solution (other ingredients of the extraction protocol were scaled up proportionately). Before the protein precipitation step, the RNaseA-treated samples were transferred into newly labelled Eppendorf tubes, leaving behind the pieces of filter paper. All the other procedures for DNA extraction were as described above.

For the brush swab, 800 µL of lysis solution was added to fully immerse the brush and the rest of the chemicals were scaled up based on the volume of lysis solution. Either the brush was removed or RNaseA-treated solutions were transferred into newly labelled Eppendorf tubes (based on the experimental design – see Chapter 3) before treating with 5 M NaCl. All the other procedures were as described above.

Table 2.2 Preparation of SSTNE DNA extraction buffer.

Reagent	Final Conc.	Amount for one litre
NaCl	0.3 M	17.5 g
Tris Base	50 mM	6.05 g
EDTA	0.2 mM	74.448 mg
EGTA	0.2 mM	76 mg
Spermidine	0.5 mM	72 mg
Spermine	0.25 mM	52 mg

Note: ddH_2O was added to make a final volume of 1 L. The solution was autoclaved and stored at 4°C. pH was c. 9.5 -10.0 without adjustment. All the chemicals were supplied by Sigma Aldrich, UK.

2.5.2 HotSHOT method

Genomic DNA was extracted using hot sodium hydroxide and Tris. A piece of fin tissue (size approximately 0.2 cm × 0.2 cm) was placed on a tissue paper to remove the ethanol and was taken into 0.2 mL PCR tube. The 50 µL alkaline lysis reagent was added to the

tube (**Table 2.3**) and incubated at 95°C for 30 min in a thermocycler. The solutions were cooled to 4°C and 50 µL neutralisation buffer (**Table 2.3**) was added to neutralise the lysis reagent. The solutions were mixed by inverting and centrifuged at 17,000 × G for 5 min; the DNA was stored at 4°C for future use.

For brush swab, a small portion of brush (mucus) was cut into small pieces and 200 µL alkaline lysis reagents was used and neutralised by adding 200 µL of neutralisation buffer. The rest of the procedures were same as described above.

Table 2.3 Solutions for HotSHOT genomic DNA extraction.

Alkaline lysis reagent			**Neutralisation buffer**		
Reagent	**Final Conc.**	**Amount for 200 mL**	**Reagent**	**Final Conc.**	**Amount for 200 mL**
NaOH	25 mM	200 mg	Tris-HCl	40 mM	1.3 g
EDTA	0.2 mM	14.88 mg			
Note: ddH_2O was added to make a final volume of 200 mL. pH of alkaline lysis reagent and neutralisation buffer was 12 and 5 respectively.					

2.5.3 DNA extraction from pooled tissue sample using SSTNE/SDS method

An equal number of male and female progeny from each family were used to make male and female progeny pools respectively (Chapter 4). An approximately equal amount of fin tissue from each individual was taken by using half of the tissue cut out with a 3 mm biopsy punch. These were pooled in a 15 mL conical centrifuge tube (**Figure 2.4**).

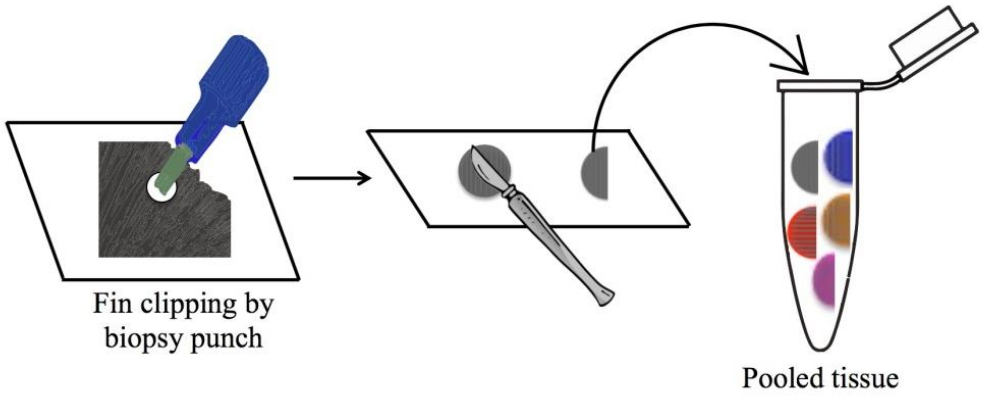


Figure 2.4 Pre-extraction pooling of fin tissue samples for DNA extraction (Original image).

After making the progeny pool (with 30 individuals), 2.5 mL SSTNE and SDS mixed buffer (lysis solution) and 50 µL proteinase K (10 mg/mL, Sigma Aldrich, UK) were added, the lid was tightly sealed with Parafilm and mixed by inverting. Pooled tissues were incubated overnight at 55^{o}C on a rotator. The digested samples were allowed to cool to room temperature (22^{o}C). Then 600 µL aliquots were transferred into newly labelled Eppendorf tubes (discarding the rest of the digestion solution). For RNaseA treatment, 30 µL (2 mg/mL) of DNase free RNaseA (Sigma Aldrich, UK) was added, mixed and incubated at 37^{o}C for an hour on a rotator. Solutions were cooled down to room temperature and 0.7 volumes of 5 M NaCl was added to precipitate protein, mixed by vortexing at high speed for 30 seconds and left on ice for 10 seconds. Centrifugation at 21,000 × G for 10 min was carried out to precipitate protein, which formed a white pellet at the bottom of the tube. Following protein precipitation, 400 µL of supernatant was pipetted out into a newly labelled Eppendorf tube without touching the protein pellet. To precipitate DNA, 880 µL of 100 % ethanol (2.20 × vol.) was added to the supernatant, mixed by 5-6 sharp inversions and at this stage thread like DNA structure should be visible. The rest of the procedures

were same as described in the section 2.5.1. The volume of each chemical used in this method was scaled down according to the number of individuals used to make each pool.

2.6 Genomic DNA quantification and DNA quality assessment

Quality and quantity of genomic DNA was measured first using a Nanodrop (ND-1000) spectrophotometer (Labtech International Ltd, UK), using the natural absorbance of light at 260 nm (for DNA and RNA) or 280 nm (for proteins).

Nanodrop spectrophotometric quantification measures nucleic acid concentration along with impurities, which might have adverse effects for subsequent analyses. It was also necessary to know the actual concentration of double-stranded DNA (dsDNA) for each sample, particularly for ddRADseq library preparation (to minimise variation in concentration among samples at the point of pooling). Thus another technique, Qubit® Fluorometric Quantitation (Life Technologies Ltd, UK) was also used for quantification of dsDNA only, which is suitable for NGS (Simbolo *et al.* 2013). It measures the concentration of nucleic acids and proteins in a sample upon binding of fluorescent dye to the target (DNA, RNA or protein). This method uses specific kits for DNA, RNA or protein and it is possible to measure only dsDNA using Qubit® dsDNA BR (Broad Range) assay. This method is very sensitive to light, temperature, human touch and quantification has to be done under dark conditions. Qubit® dsDNA BR assay requires two standards (we used 0 and 100 ng/µL). The specific tubes (0.5 mL) for Qubit quantification were used and labelling was done on top of the tube for standards and sample tubes. Qubit® working solution was prepared by diluting Qubit® dsDNA BR Reagent in Qubit® dsDNA BR Buffer (1:200), mixed them slowly upside down using another black microcentrifuge tube rack on top (without touching the tube), centrifuging briefly then leaving at room temperature for a few seconds. The 95

µL of Qubit® working solution was added to each of the tubes. An aliquot of 5 µL of each Qubit® standard was added into each standard tube and 1 µL of genomic DNA (100 ng/µL) was added to each sample tube. TE buffer (4 µL) was also added to each sample tube to give a final volume of 100 µL in all tubes. Tubes were always kept in a black microcentrifuge tube rack, mixed very slowly upside down followed by very brief centrifugation and holding at room temperature for at least 2 min. Standards were measured first and then the rest of the samples were measured. The Qubit® Fluorometer calculated the concentration automatically based on the amount of fluorescence signal which was directly proportional to the concentration of the DNA in the solution.

The samples used for ddRADseq were quantified using both techniques but the ones used for standard genotyping were quantified only by Nanodrop spectrophotometer. The range of Qubit/Nanodrop ratio was 0.6 – 0.9. After quantification, each sample was diluted with 5 mM Tris to the concentration required for each method used, i.e. ~50 ng/µL for capillary sequencing and KASP assay and 8.0 ng/µL for ddRADseq.

The quality and quantity of DNA were also assessed by 1 % agarose gel electrophoresis. Agarose gel was prepared in 0.5× TAE and an aliquot of 0.5 µL (5 mg/mL) EtBr was added during gel preparation to allow visualization of DNA under UV light. Approximately 50 ng genomic DNA was loaded on the gel and run at 70-100 V. DNA quantity on the gel was assessed by comparison to the known size standard (ladder).

2.7 Double-digest restriction-site associated DNA (ddRAD) library preparation and sequencing

The protocol used for ddRAD library preparation was based on the protocol described in Peterson *et al.* (2012) with some modifications. DNA quantity and quality is a major

concern for this procedure. Actual concentrations of DNA from individual and pooled samples were measured by Qubit® Fluorometric Quantitation described in section 2.6 and diluted to 8.0 ng/µL. Examples of genomic DNA (24 ng) for individual and pooled samples used in this study are shown in **Figure 2.5**.

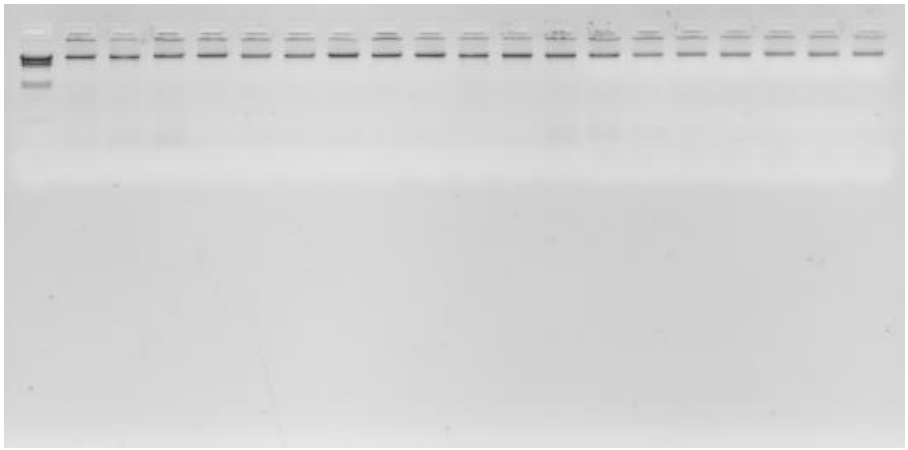


Figure 2.5 Consistency of genomic DNA (24 ng) for samples used in BSA-ddRAD sequencing after Qubit quantification, with no visible sign of RNA contamination.

In total three ddRAD libraries were constructed. The pooled tissue was considered as a single individual during the library preparation. The protocol used for ddRAD library preparation and sequencing is described step by step below:

2.7.1 Restriction enzyme digestion

Genomic DNA was subjected to digestion by *Sbf*I and *Sph*I restriction enzymes. *Sbf*I has an 8 base pairs (CCTGCA^GG) recognition sequence, i.e. is a very infrequent cutter, while *Sph*I has a 6 base pairs (GCATG^C) recognition sequence, i.e. is a more frequent cutter. Restriction digestion master mix was prepared using the required volume of *Sbf*I and *Sph*I

enzymes in CutSmart Buffer and MQ water (**Table 2.4**) and kept on ice. DNA samples (24 ng for each sample) were plated out into a 96 well PCR plate supported by a plastic rack. Aliquots of 3 µL of master mix were added to the individual DNA samples and mixed by pipetting three times, avoiding bubbles. The PCR plate was sealed using an adhesive sealing sheet (Thermo Scientific, UK) followed by gentle vortexing and brief centrifugation (twice) to mix properly for better digestion. Samples were then incubated at 37^{o}C for 90 min in a thermal cycler with the lid temperature at 42^{o}C. Following incubation, samples were cooled at room temperature (22^{o}C) for 5 min on the bench.

2.7.2 Ligation of adapters

During the restriction enzyme digestion, pre-prepared and labelled adapter/barcode mixes were removed from the freezer and thawed in a fridge. After thawing, barcode mixes were kept on ice until use. Each mix contains individual-specific P1 and P2 adapters (combination of 5 and 7 base pairs long barcodes). The P1 adapter is compatible with *Sbf*I sticky ends and P2 is with *Sph*I. Aliquots of 3 µL barcode mix were added to the restriction enzyme digested fragments using filter tips and mixed up and down three times avoiding bubbles. It is important to record which adapter mix is used for which sample to identify the samples during data analysis. Digested fragments with adapter mix were incubated for 10 min at room temperature on the bench to allow initial annealing of sticky ends. During incubation ligase master mix was prepared using the required volume of T4 Ligase and rATP in CutSmart Buffer and MQ water (**Table 2.4**). 3 µL of ligase master mix was added to each sample and mixed by pipetting. The PCR plate was sealed followed by vortexing and briefly centrifuging (twice) to ensure mixing. Ligation was carried out at 22^{o}C for two and half hours.

Table 2.4 Chemicals required for restriction digestion and adapter ligation in ddRAD library preparation.

DNA digestion mix			**Barcode and ligase mix**		
Components	**Final conc.**	**Vol. (µL)**	**Components**	**Final conc.**	**Vol. (µL)**
DNA		3.000 (24 ng)	Barcode mix (*SbfI*:*SphI* 1:12)		3.000
10× CutSmart Buffer	1×	0.600	10× CutSmart Buffer	1×	0.300
SbfI Enzyme (20 U/µL)	20 U/µg	0.024	rATP (100 mM)	1 mM	0.120
SphI Enzyme (20 U/µL)	20 U/µg	0.024	T4 Ligase 2 M ceU/mL	2K ceU/ug	0.024
Water		2.352	Water		2.556
Total		6			6

2.7.3 Multiplexing of samples and purification

The ligation reaction was stopped by addition of 2.5 volumes of PB buffer (MinElute PCR purification kit, Qiagen, UK) to each sample. PB buffer has high salt concentration and low pH (yellow colour as pH indicator), which helps DNA to bind to the membrane of the column during purification. All the samples were multiplexed into one Bijou tube (7 mL, Thermo Scientific, UK) by individual pipetting, ensuring collection of all the liquid. Pooled samples were inverted to mix properly and left on the bench to settle down. If the colour of the solution turned to violet, about 2-3 µL 3 M Sodium Acetate was added and shaken very slowly to get light yellow coloured solution to adjust the pH (pH 5.2). Solutions were loaded onto a single column in sequential aliquots (550 µL each time) and centrifuged at 17,800 × G (same force was used throughout the purification step) for 10 sec with the lid closed between aliquots. Each time the column was placed into a new clean collection tube. After the final aliquot had been added to the column, it was centrifuged for 1 min with the lid open (help remaining impurities to flow through the column). The column membrane was thoroughly washed once with ethanol-containing wash buffer (aliquots of 720 µL) to

remove any remaining impurities attached to the column and centrifuged for 1 min with the lid closed. The column was then transferred into a new collection tube followed by centrifuging for 1 min with the lid open. The washed column was kept in a nuclease-free 1.5 mL Eppendorf tube and was incubated in a heat block at 60^{o}C for 5 min with the lid open to remove excess ethanol. An aliquot of 35 µL of warmed elution buffer was added to the column, which was left in the tube at room temperature for 3-4 min, followed by 30 sec centrifugation. Another 35 µL warmed elution buffer was added again to the column, which was left at room temperature for 3-4 min and then given a final centrifugation for 1 min. The 64 µL library solution was stored on ice overnight.

2.7.4 Size selection on agarose gel

Size selection was done on a 1.1 % agarose gel. For the gel preparation, 500 mL 0.5× TAE buffer was prepared on the first day of library preparation and kept in the fridge. The 0.42 g agarose powder was added to 38 mL 0.5× TAE buffer to make the 1.1 % agarose gel. The agarose powder and buffer mixture and the gel apparatus were also kept in the fridge overnight.

On the second day, the agarose gel was heated up, poured on the gel cast tray (without addition of EtBr) and kept in the fridge to set and cool down. The sample loading well was 25 mm wide and 1.5 mm thick while the flanking wells for molecular weight markers were 3 mm wide with the same thickness. Once the cooled gel set it was transferred to the gel chamber and cooled buffer (from the fridge) was poured into the gel chamber to submerge the gel. Electrophoresis was done in a cooled condition (gel chamber was surrounded with ice). The idea behind the chilled electrophoresis is to minimise the diffusion of small fragments in the gel and to get more precise size selection. The lid was closed and the gel was run for 10 min at 40 V to ensure electrical connections were good. The purified,

concentrated 64 µL ligated DNA sample was prepared by adding 13 µL of 6× loading buffer (including Bromophenol blue, BPB), mixed by flicking, brief centrifuged and kept on ice. Each of the two molecular weight ladders was processed by mixing 4 µL of 416/706 bp marker with 3.6 µL of 6× loading buffer and 12.4 µL of elution buffer. First the molecular weight ladders were loaded into the flanking wells to make sure the wells are sound. Then the entire sample (77 µL) was loaded carefully using a 20 µL pipette and left for a few minutes to settle down in the well. The gel was run at 40 V for 4 min, 65 V for 5 min and finally at 90 V for about an hour. The electrophoresis was stopped when the BPB band had migrated 3.5 cm from the origin. A 2 cm wide and 3 cm long section, with the BPB band in the middle, was cut using a clean scalpel to ensure the desired size selection, leaving 2 mm on either side to avoid electrophoretic side effects. The gel was cut at an angle to remember the orientation of the gel (**Figure 2.6**). The excised gel was kept on a clean tray with a lid on top to avoid contamination and stored in a fridge. The remainder of the gel was stained with EtBr (5 µL of 5 mg/mL EtBr) in 0.5× TAE for 10 min to allow visualization of the marker bands under UV light. A small notch was cut to mark the molecular weight markers. The stained gel was rinsed in water for 5 min. The excised gel was removed from the fridge and placed in the right position in the washed remainder of the gel. The desired gel section was cut carefully with a clean scalpel according to the positions of the marker bands (to retain the portion containing 400–700 bp fragments). The gel slice was weighed and divided into three parts (0.24 g each), each of which was placed into a 1.7 mL Eppendorf tube (**Figure 2.6**). After size selection, the rest of the gel was stained with EtBr solution and a photograph taken.

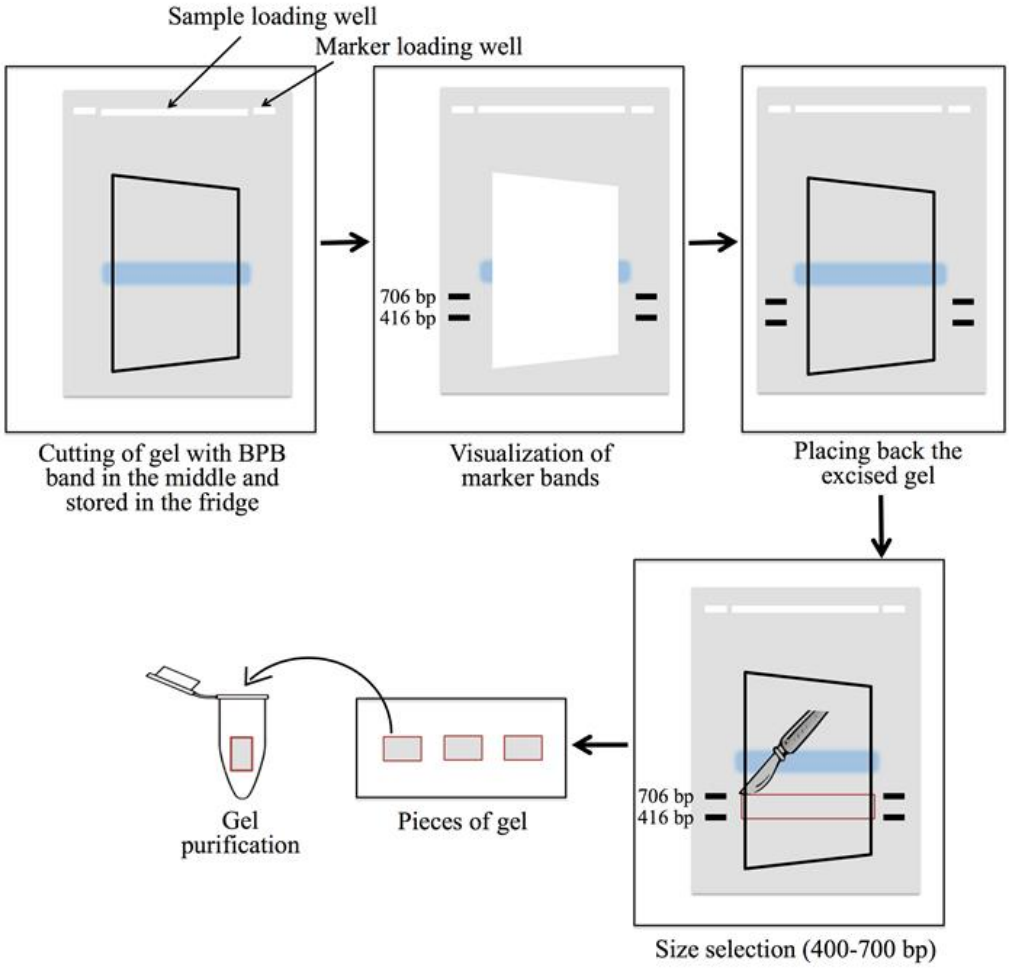


Figure 2.6 Size selection of the ligated DNA fragments on the 1.1 % agarose gel (Original image).

2.7.5 Purification of the gel

The gel was dissolved in 3 volumes (720 µL for 0.24 g) of QG buffer (MinElute gel purification kit, Qiagen UK), using a paddle mixer for 10 min to help this. After dissolving, samples were centrifuged briefly and one volume (240 µL for 0.24 g) of fresh isopropanol was added and mixed thoroughly, followed by brief centrifugation. The QG buffer/gel mix was loaded on a single MinElute column in sequential aliquots (no more than 550 µL volume each time) and centrifuged for 10 sec at 17,800 × G (same force was used throughout the purification step) for each aliquot, and column was loaded into a new clean

collection tube for each aliquot. After final aliquot, centrifuged for 1 min and an additional aliquot of 500 µL of QG buffer was also added to remove impurities and centrifuged for 1 min with the lid closed. To give a final wash, 720 µL ethanol containing wash buffer was added and centrifuged for 1 min. The column was transferred to a new collection tube and centrifuged for 1 min with the lid open. The column was placed into a 1.5 mL Eppendorf tube and incubated for 5 min at 60^{o}C to remove excess ethanol. Samples were eluted with 2 × 35 µL elution buffer and finally, 66 µL of library template was obtained.

2.7.6 Test PCR and bulk PCR amplification

A test PCR was done to select the appropriate number of PCR cycles for the bulk PCR to get an adequate amount of amplified library for the sequencer. The PCR was performed with NTC (no template control), 0.5 µL, 1 µL and 2 µL of library template. The PCR master mix was prepared by adding 25 µL of 2× Q5 Hot-start High Fidelity DNA polymerase (NEB), 0.8 µL of 10 µM Illumina specific primer mix and 16.2 µL of MQ water. The 10 µM primer mix was prepared by using 20 µL of 100 µM P1 primer, 20 µL of 100 µM P2 primer and 160 µL of MQ water. Four PCR tubes were labelled and 10.5 µL PCR master mix was added into each tube. The NTC, 0.5 µL, 1 µL and 2 µL of library templates was added to the respective labelled tube and topped up with MQ water to 12.5 µL. At first 16 cycles of PCR were used for NTC and 0.5 µL template, whereas 13 cycles were used for 1 µL and 2 µL templates. The PCR conditions were: initial denaturation at 98^{o}C for 40 sec, 16/13 cycles of PCR where the denaturation at 98^{o}C for 10 sec, annealing at 65^{o}C for 30 sec and extension at 72^{o}C for 40 sec, with the final extension at 72^{o}C for 2 min. Amplified products were loaded on a 1.5 % agarose gel (in 0.5× TAE). A 5 uL aliquot of the amplicon was diluted with 3 µL 3× loading buffer and the molecular weight ladder was prepared with 0.7 µL of 100 bp marker in 7.5 µL 1× loading buffer. The 1 µL library

template (with 7.5 µL 1× loading buffer) was also run on the same agarose gel to see the intensity of amplification. Samples were run until the BPB band moved to 2.5 cm from the origin (**Figure 2.7**).

Based on the first gel run, 1 µL of template in 12.5 µL reaction volume and 11 cycles of PCR were selected for bulk PCR. To reduce the PCR bias, a single master mix was prepared using 32 µL templates to produce about 400 µL of amplified library and aliquoted (12.5 µL reaction volumes) into individual wells in a 96 well PCR plate, followed by running with the same PCR conditions with 11 cycles. Following amplification, products were pooled into one single tube and 5 µL of bulked product was run on a 1.5 % agarose gel to confirm the amplification (**Figure 2.7**).

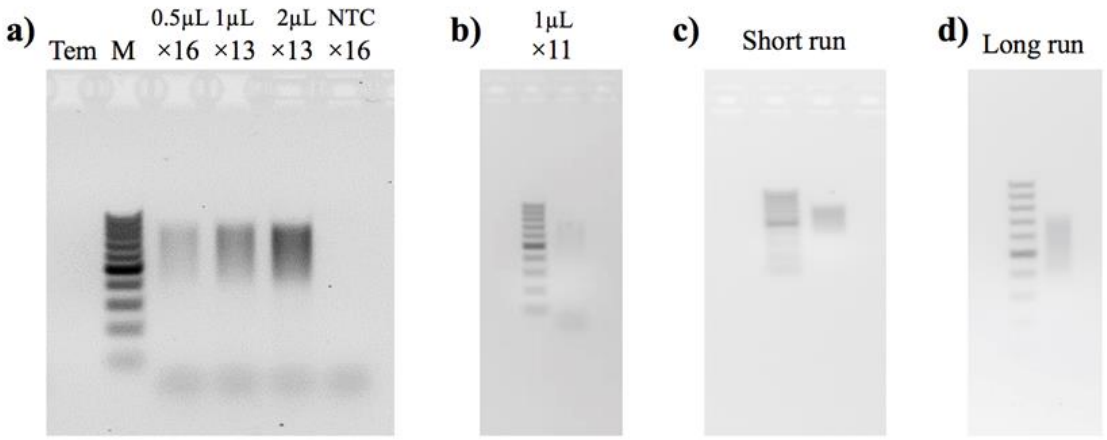


Figure 2.7 Amplified library on the 1.5 % agarose gel; a) Test PCR (in 13 and 16 cycles with 0.5 µL, 1 µL and 2 µL library templates plus NTC) products, b) Bulk PCR products at 11 cycles with 1 µL library templates, c) Short gel run of the purified amplified library and evidence of no lower fragments or primer dimers, d) Long gel run of the purified library and evidence of accurate size selection (400-700 bp), Ladder - 100 bp.

2.7.7 Purification of the amplified library

Amplified libraries were purified twice, i.e. column purification followed by a paramagnetic bead clean up (AMPure XP, Beckman Coulter, UK). Column purification was as described in section 2.7.3 followed by eluting into 2 × 28 µL elution buffer and 50.5 µL of amplified library was obtained from the first wash.

For magnetic bead clean up, Ampure beads were removed from the fridge and kept at room temperature to equilibrate. Beads were mixed carefully and an equal volume of beads (50.5 µL) was added to the 50.5 µL amplified library followed by mixing with a pipette, ensuring that the solutions were at the bottom. The sample was left for 5 min at room temperature, which will help to bind DNA fragments to paramagnetic beads, then the lid was opened carefully and the tube placed in a magnetic stand. The magnetic stand was kept undisturbed for about 3-4 min until the impurities was completely separated from the beads. The supernatant was pipetted out carefully from the tube placed in a magnetic stand and discarded. An aliquot of 190 µL 73 % ethanol was added to the tube to give a first wash and left in the magnetic stand for 30-60 sec. Following pipetting out all the ethanol, a second wash was given again with the 73 % ethanol to remove all the contaminants and all the ethanol was pipetted out from the tube. After the second wash, the tube was removed from the magnetic stand and kept in a heat block at 60^{o}C for 2-3 min with lid open to remove all the excess ethanol. The washed beads were re-suspended in 20 µL warmed elution buffer and mixed by gentle pipetting, followed by incubation at 60^{o}C for 2-3 min with lid closed. Following incubation, the tube was placed again in the magnetic stand with the lid open (for the collection of purified library without touching the magnetic beads) and left for 3-4 min without disturbance until the purified DNA fragments have completely separated from the beads. The supernatant (contained purified amplicon) was pipetted off carefully into a

newly labelled Eppendorf tube without touching the beads, 14.5 µL of amplicon library was obtained.

2.7.8 Final quality check and quantification of the library

A fraction (1 µL) of the amplicon library was run on a 1.5 % agarose gel to estimate the quality and size range of the library. One photograph was taken when the BPB band had migrated about 1.5 cm from the origin and a second one was taken when the BPB band was about 3 cm away from the origin. A clear smear was evident (about 400 – 700 bp long) on the gel and there was no indication of low molecular weight DNA (**Figure 2.7**).

For the quantification of the amplicon library, the Qubit® dsDNA HS (High Sensitivity) assay was used. The working solution was prepared by diluting dsDNA HS Reagent in dsDNA HS Buffer (1:200) and the final volume for each sample was 200 µL. Two standards (0 and 10 ng/µL, 10 µL volume from each standard) were used and 190 µL of working solution was added to each standard. The amplicon library (1 µL) and template library (1 µL) were also quantified in duplicate, where 190 µL of working solution was used in each and made up to 200 µL with TE buffer. The quantification procedure was as described in section 2.6. The average concentration of the amplicon library and template library were 11.2 and 0.458 ng/µL respectively.

2.7.9 Preparation for library clustering and sequencing

The purified amplicon library was normalized to 10 nM stock in EB buffer and 1 % Tween 20. The stock amplicon library (10 nM) was diluted to 4 nM (2 µl of 10 nM library and 3 µl ddH_2O) followed by denaturation in 5 µl of fresh 0.2 M NaOH at room temperature for 5 min. The denatured library was diluted again to 20 pM with 990 µl pre-chilled HT1 (hybridization) buffer. PhiX was used as an internal control during sequencing and was

diluted to the same loading concentration (20 pM) as the amplicon library. The final concentration of the loading library was 10.2 pM; of which 5 % was PhiX (290.7 µl 20 pM amplicon library, 15.3 µl 20 pM PhiX library and 294.0 µl HT1 buffer). The final library was heat denatured at 98^{o}C for 2 min followed by cooling in an ice bath for 5 min before loading on an Illumina MiSeq cartridge. The flow cell was loaded onto the MiSeq machine according to the manufacturer's instruction. The v2 sequencing chemistry was used and the library was run in a single lane for 300 cycles for 161 bases paired-end reads. The general workflow of the ddRADseq is presented in **Figure 2.8**.

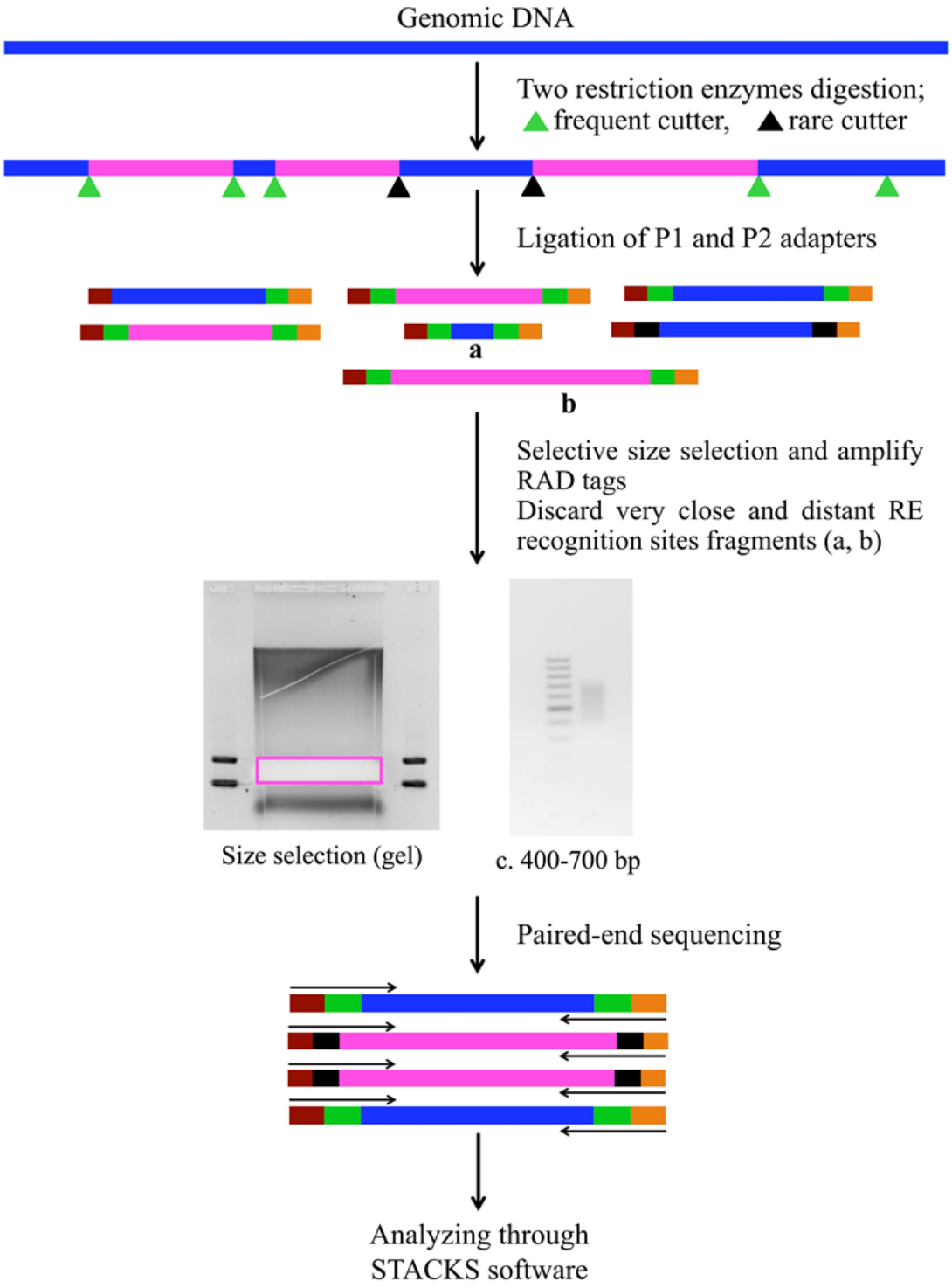


Figure 2.8 General principle of ddRADseq (Original image, Peterson *et al.* 2012).

2.8 Genotyping assays

2.8.1 Microsatellite marker genotyping

Genomic DNA samples were amplified by PCR. In order to reduce the cost and to increase the genotyping output, the fluorescent labelled tailed primer method originally developed by Boutin-Ganache *et al.* (2001) was used. Genomic DNA amplified during PCR was labelled with one of three fluorescent dyes that allow detection of the fragments by the sequencer, via a complementary DNA "tail" (**Table 2.5**). The tail was attached at the 5´ end of either the forward or reverse primer. The GTTT 'tail' was also added to the 5' end of unlabeled primers to enhance extra A addition (Tonteri *et al*. 2008). The general principle for the tailed primer method is shown in **Figure 2.9**.

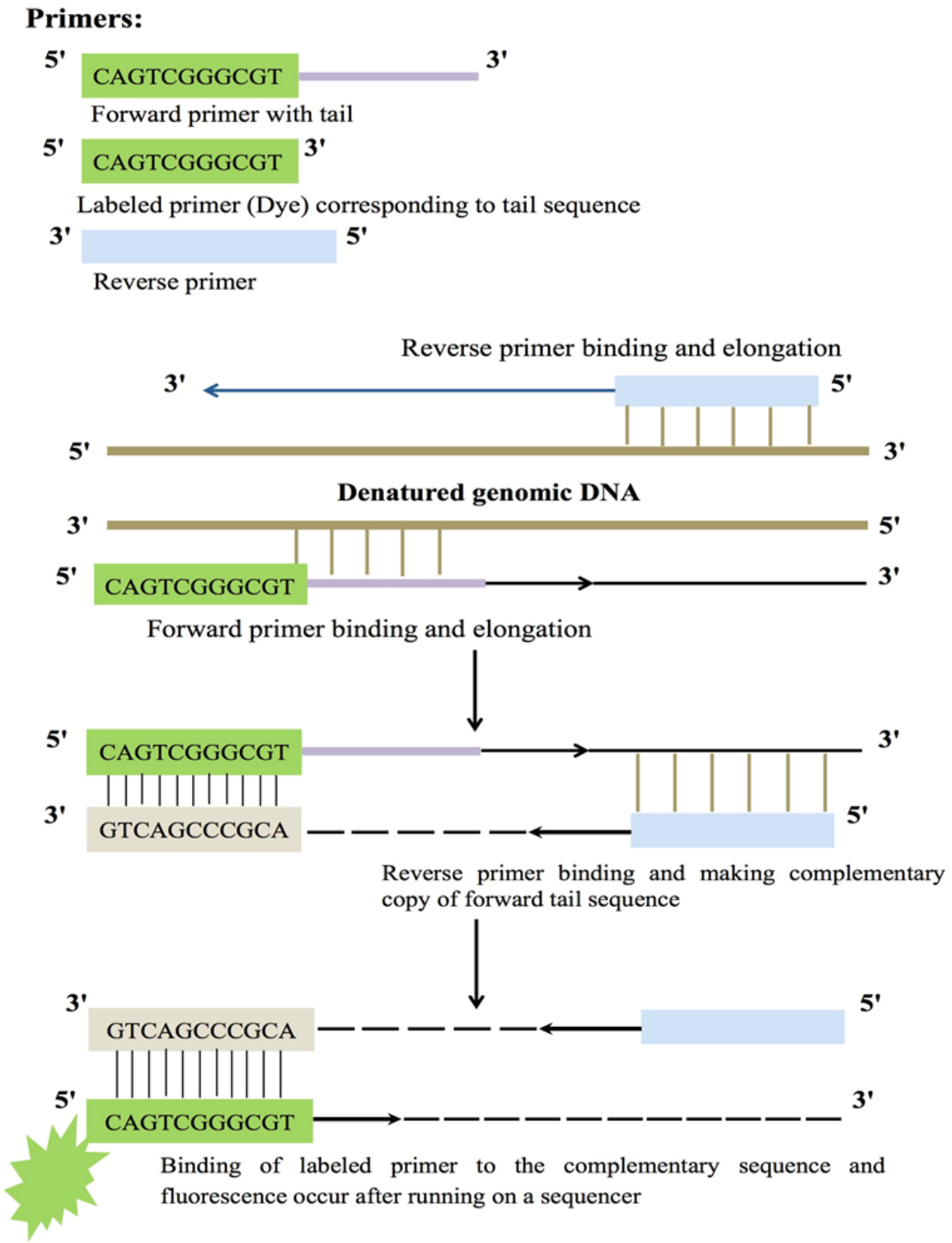


Figure 2.9 Working principles of tailed primer method (Original image, Boutin-Ganache *et al.* 2001).

Ready to use Promega master mix (Promega, UK) was used for PCR. This master mix contains buffer, Taq DNA polymerase, dNTPs and $MgCl_2$. Primers were supplied by MWG Eurofins (Germany). Genomic DNA was amplified with the labelled and unlabelled primers during the first round of PCR that generated DNA with the tail sequence. The complementary sequence of the tailed primer was produced by the second cycle of PCR. In the next PCR cycle, the dye-labelled primer attached to the 3´ end of the denatured DNA containing the complementary sequence of the tailed primer. At this stage genomic DNA was amplified containing the fluorescent dye, which was detected by the sequencer. The PCR master mix and reaction cycles are given in **Table 2.6** and **2.7**.

Table 2.5 Oligo sequences of fluorescent labeled primers.

Oligo names	**Dye**	**Oligo sequences**	**Size**	**GC Content**
Godde_Black	D2	5' CATCGCTGATTCGCACAT 3'	18	50 %
CAGtag_Green	D3	5' CAGTCGGGCGTCATCA 3'	16	62.5 %
M13A_Blue	D4	5' GGATAACAATTTCACACAGG 3'	20	40 %

Table 2.6 PCR reaction volumes for fluorescent labelled amplification of desired segments by tailed primer method.

Chemical components	**Per reaction**	**5** µL **reaction**	**10** µL **reaction**	**15** µL **reaction**
2× Promega Master Mix	1×	2.5	5	7.5
10 uM Labeled primer	0.3 uM	0.15	0.3	0.45
10 uM UNH995 R Primer	0.3 uM	0.15	0.3	0.45
1 uM UNH995 F Tailed Primer	0.02 uM	0.1	0.2	0.3
DW	-	1.1	3.2	5.3
DNA	~50 ng/µL	1	1	1

Table 2.7 Thermal cycling conditions for amplification of microsatellite marker.

Step	Temperature	Time	Cycles
Initial activation	95°C	5 min	-
Denaturation	95°C	30 sec	
Annealing	60°C	30 sec	35
Extension	72°C	30 sec	
Final extension	72°C	10 min	-
Pause	10°C	30 sec	-

2.8.1.1 Capillary sequencing and allele annotation

The labelled PCR fragments were genotyped on a CEQ™ 8800 capillary sequencer (Beckman Coulter®, USA). Before running on a sequencer, labelled PCR fragments were run on 1.5 % agarose gel to ensure the presence of the amplified fragments and to assess the volume of the PCR products to be used for sequencing. Multiplex genotyping was performed by using PCR fragments attached with three different dyes and run in a single well (96 well sample plate) to reduce the genotyping cost. Master mix was prepared using 30 µL SLS (Sample Loading Solution) and 0.4 µL DNA size standard kit-400 (SS400, Beckman Coulter®, USA). The SS400 contains fragments labelled with D1 (red dye) and was used as a molecular weight ladder to identify the fragment sizes on the sequencer. Aliquots of 29 µL of master mix were placed in wells of a 96 well sample plate (Beckman Coulter®, USA) and PCR products (based on the gel intensity 2.5 µL, 3.5 µL and 2.5-3 µL for green, black and blue labelled PCR fragments respectively) were added and mixed thoroughly by pipetting. One drop of mineral oil was also added on top and the plate was centrifuged. An electrophoresis buffer plate (96 well, Beckman Coulter® USA) was also prepared. Beckman Frag3-45 genotyping method was followed to run the samples on the sequencer.

Raw data were generated from the machine and checked first using the default parameter of the CEQ software designed for the Beckman-Coulter sequencer. This software identified and quantified the detected allelic fragments for the ladder and unknown samples. Fragment analysis parameters for the three different dyes were set manually within the software and afterwards the software binned the allele sizes automatically. The fragments for the molecular weight ladder (D1, red dye) were checked manually and if the ladder migrated properly then the fragment size for unknown samples was measured based on the ladder and checked each of the fragments manually. The samples which had migration problems were repeated. All the analysed data were transferred to an Excel spread sheet for further analysis.

2.8.2 Single nucleotide polymorphic marker genotyping

Single nucleotide polymorphisms **(**SNPs**)** are one of the most common types of genetic variations in plant and animal genomes and were used as molecular markers for QTL analysis and association studies. Kompetitive Allele Specific PCR (KASP: LGC Genomics, UK) genotyping system is one of the simplest, most economical ways to genotype individual SNPs.

KASP genotyping is based on the competition between two allele-specific primers to bind to DNA during PCR and detection of fluorescent signals based on the individual's genotype. KASP genotyping reaction reagents are composed of KASP assay (primer mix), KASP master mix and genomic DNA. The KASP assay contains three primers, two allele-specific forward primers ending at the SNP and one common reverse primer. Each forward primer has a unique tail sequence, which corresponds to one of the two dye (FAM or HEX with Quencher) sequence. The KASP master mix contains FRET (fluorescence resonant energy transfer) cassette (dyes), reference dye, Taq DNA polymerase, dNTPs and $MgCl_2$ in

a buffer solution. The principle behind this KASP genotyping system is shown in **Figure 2.10**.

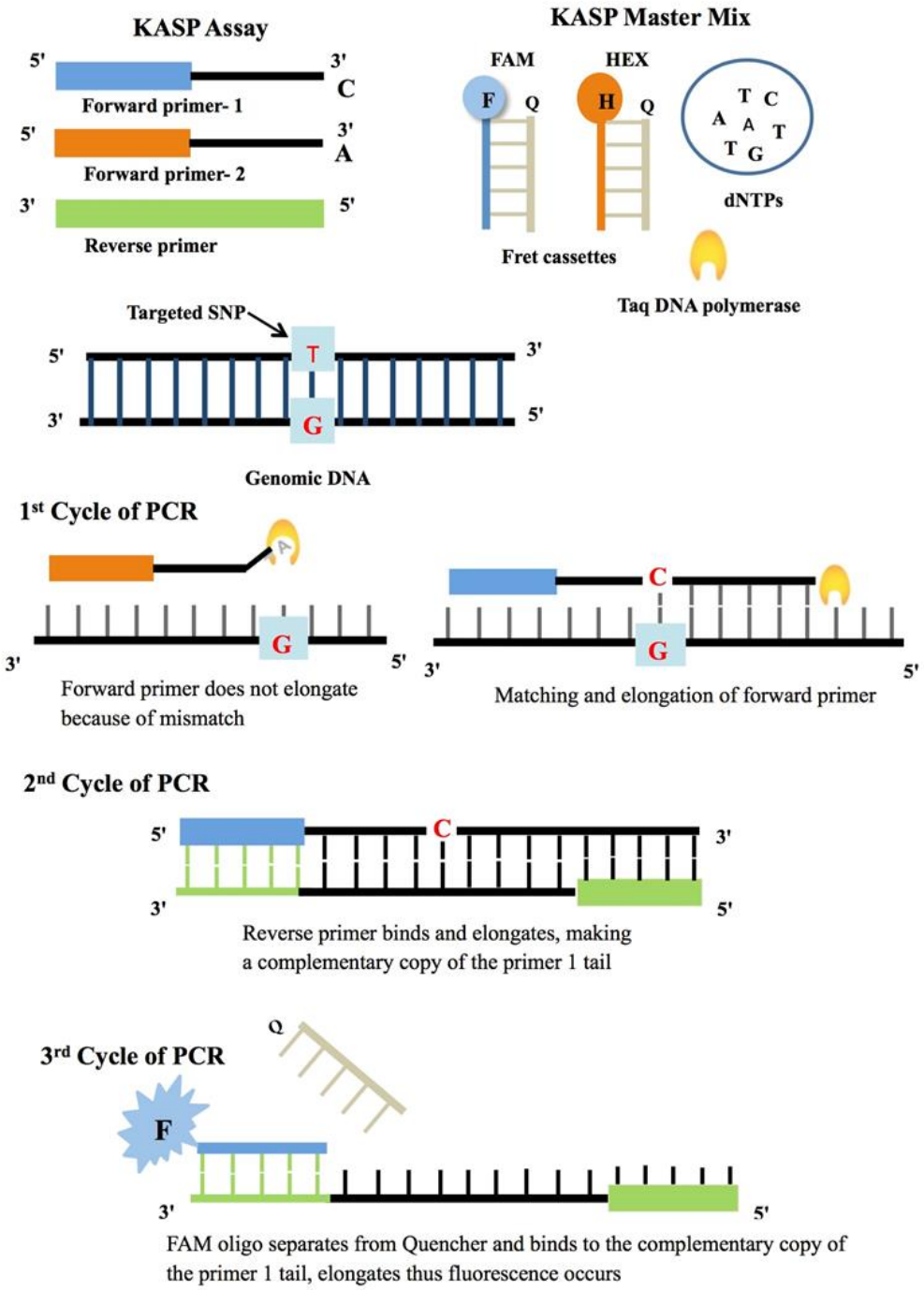


Figure 2.10 Mechanism of KASP genotyping chemistry (Original image, Semagn *et al.* 2014).

The KASP master mix and allele specific primers (KASP assay) were provided by LGC Genomics based on the supplied marker and flanking sequences (from ddRAD loci). Both dry and wet DNA methods were followed in KASP genotyping system. For the dry method, DNA samples were pipetted in wells in 96-well plates, then left in a clean fume cupboard for about an hour until they had dried up. They were then stored in a fridge for future use. Chemical reaction volumes and thermocycler conditions for PCR are given in **Table 2.8** and **2.9**.

Table 2.8 PCR reaction volumes (10 µL) for dry and wet DNA methods for KASP genotyping.

Components	**Wet DNA method (µL)**	**Dry DNA method (µL)**
KASP master mix	5	5
KASP Assay mix	0.14	0.14
ddH_20	3.86	4.86
Genomic DNA (50 ± 5 ng/µL)	1	1

Table 2.9 PCR thermal cycling conditions for KASP genotyping.

Step	**Temperature**	**Time**	**Cycles**
Initial activation	94^{o}C	15 min	-
Denaturation	94^{o}C	20 sec	10
Annealing and elongation	65-57^{o}C (dropping 0.8^{o}C per cycle)	1 min	
Denaturation	94^{o}C	20 sec	34
Annealing and elongation	57^{o}C	1 min	

Following PCR, plates were run on a Techne Quantica® machine (Barloworld Scientific Ltd UK), which detects the fluorescent signals from the labelled PCR products, and used an in-built software (Quansoft, version 1.1.21) developed for that machine to analyse data. If an individual is homozygous at a given SNP, the fluorescent signals will be from only one of the two dyes (corresponding to one of the two possible alleles), while in case of a heterozygote both fluorescent signals will be detected. **Figure 2.11** illustrates an example of this graphically. No template controls (NTC) were also used to detect the cross contamination.

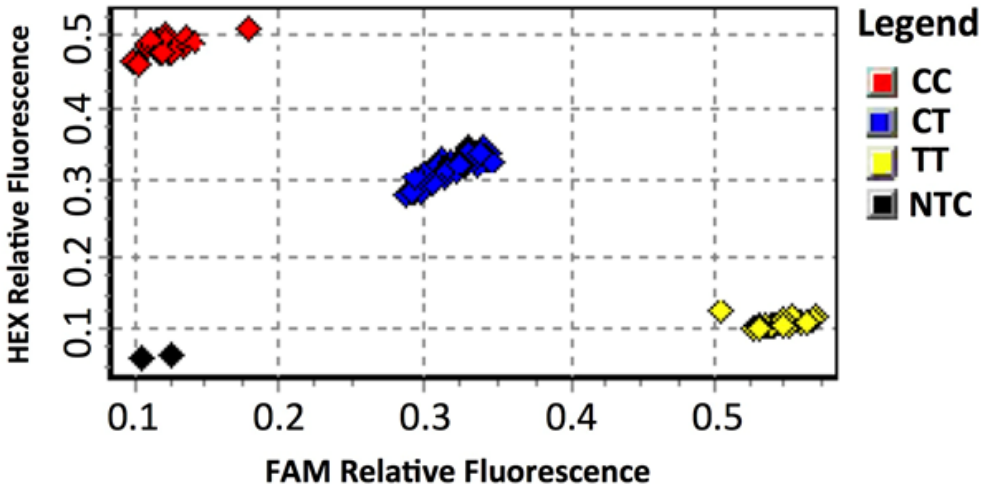


Figure 2.11 Allelic distribution graph through KASP genotyping system. Red and yellow symbols show homozygotes for C and T allele respectively; blue symbols show heterozygotes (CT); black symbols show no template controls (NTC).

2.9 Data analysis

2.9.1 Computational methods for generating ddRAD loci

Following sequencing, two fastq files were generated from the sequencer for paired-end sequencing and the quality of the sequencing was checked using FASTQC software (Version 0.11.2). Raw reads were processed using process_radtags in the STACKS (version 1.27) and low quality reads (score under 20), reads missing the restriction enzyme site or with ambiguous or unpaired barcodes were filtered out. Retained reads were sorted into RAD loci using both reference-based (published tilapia genome, Broad Institute of MIT and Harvard genome assembly Orenil1.1; Brawand *et al.* 2014) and *de novo* assembly approaches (employing ref_map.pl and denovo_map.pl components respectively). Detail analysis for each experiment is described in each chapter.

2.9.2 General statistics and association analysis

An association analysis between SNP genotype and phenotypic sex was conducted for each family and broodstocks using SNPassoc package in R (version 3.1.3). A generalized linear model was applied under the function *WGassociation* to test the magnitude of association between SNP genotypes and phenotypic sex. Significant *p*-values were corrected for multiple testing using Bonferroni correction method. Chi-square test, G-test and Fisher's exact test were also performed in R using respective packages (version 3.1.3) for association study.

Chapter 3

Mucus DNA sampling method development

Minimally invasive DNA sampling from fish mucus for standard genotyping and next generation sequencing

Taslima K.[1,2], Taggart J.B.[1], Wehner S.[1], Davie A.[1], McAndrew B.J.[1], Penman D.J.[1]

[1]Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2]Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh

Status: Two different articles were published from the results of this chapter:

Aquaculture Research (doi: 10.1111/are.12809); Conservation Genetics Resources (doi:10.1007/s12686-016-0614-z).

Contributions:

Experimental design, sample collection and preparation, fish rearing, genomic DNA extraction, genotyping of samples with microsatellite and single nucleotide polymorphic markers, sample assessment with universal bacterial primer, construction of the ddRAD library and sequencing (under the supervision of John B. Taggart), data analysis were conducted by the author of this thesis. Data compiling and first draft of the manuscript was written by the author of this thesis who was also entirely involved in the subsequent corrections of the manuscript. The other co-authors contributed in the experimental design, analysis of generated sequence from MiSeq and editing the manuscript.

3.1 Abstract

Genotypic information is valuable in the management and understanding of captive and wild population structure. Minimally invasive sampling methods reduce potential ethical and welfare issues in obtaining DNA samples. Different sampling procedures (swab brushes and filter paper), DNA extraction methods (salt precipitation and HotSHOT) and genotyping assays (microsatellite and single nucleotide polymorphism) were tested in Nile tilapia. The mucus sampling methods tested gave adequate quantities and quality of DNA for most standard genotyping methods, although high bacterial DNA amplification was found from DNA from mucus. Mucus sampling using filter paper is convenient and more flexible than swab brushes in terms of subsequent processing of samples for DNA extraction. DNA from mucus, muscle and fin samples from two Nile tilapia was used for double digest restriction site associated DNA (ddRAD) sequencing. Following sequencing, about 8,000 and 14,000 RAD loci were retrieved from reference and *de novo* genome assembly respectively from each sample. DNA from the three sources showed similarity in terms of the numbers of polymorphic and bi-allelic loci. The > 7.5 K (reference-based genome assembly) and > 13 K (*de novo* genome assembly) RAD loci were common to all three samples in both fish, with no evidence of bacterial contamination in the mucus sequences. NGS-mediated genetic analysis of mucus DNA will not only increases the potential applications of NGS but also will contribute to conservation genetics and welfare. Fish mucus samples could be used as genetic material for many types of molecular analysis instead of more invasive sampling.

Keywords: *Oreochromis niloticus*, fish mucus, DNA, genetic marker, PCR, ddRADseq.

3.2 Introduction

Genotypic information is widely used in the conservation and management of wild fish (Schwartz *et al*. 2007) and is increasingly valuable in the management of captive populations (e.g. for verification of pedigree or for marker-assisted selection; Perez-Enriquez *et al*. 1999; Sonesson 2007). DNA can be obtained from diverse sources including hair, faeces, urine, feathers, shed skin, saliva and egg shells without affecting animals. In fish, invasive DNA sampling such as fin biopsy or blood collection, scales, sperm/eggs or killing of fish, particularly in the case of small fish, has been common practice (Cummings and Thorgaard 1994; Turtinen and Juran 1998; Wasko *et al*. 2003; Campanella and Smalley 2006; Mirimin *et al*. 2011). Tissue biopsy may have negative impacts on fish, potentially including infection, alters individuals phenotype and effects on survival, growth or behaviour which could also influence the behavioural research (Le Vin *et al*. 2011); for example, the adipose fin, commonly removed from salmonids for marking and DNA extraction, is now known to show sexual dimorphism (Hisar *et al*. 2013) and is thought to function as a flow sensor (Buckland-Nicks *et al*. 2012).

As an alternative to invasive or destructive sampling, DNA samples can be collected in a non-invasive way (source of the DNA is left behind the animal, for example hair, faeces, shed skin or feathers, egg shells etc.) or minimally invasive way (without biopsy and giving the animal as little stress as possible). Potential drawbacks of such sampling methods were thought to include low DNA quantity or poor quality/degraded DNA (e.g. which could impede amplification of long sequences), the presence of PCR inhibitors, and allelic dropout/genotyping errors (Taberlet *et al*. 1999; Lieber *et al*. 2013). The quality of the DNA is also a major concern in studies on population genetics and genetic diversity of threatened and endangered species. In case of fish living in a high density in the wild (during spawning

aggregation which is the best time/location to take samples for population genetics studies) and culture system, they may rub against each other, cause cells to slough off and could cause contamination of DNA samples from individuals, leading to errors in genotyping and sequencing results following mucus sampling. On the other hand, the potential of minimally invasive or non-invasive DNA sampling has become increasingly recognized, and this has led to recent publications investigating extraction of DNA from mucus taken from the skin or mouth cavity (Campanella and Smalley 2006; Livia *et al*. 2006; Le Vin *et al*. 2011; Mirimin *et al*. 2011; Reid *et al*. 2012; de Verdal *et al*. 2014a).

Mucus swabbing from fish has proved to be simple, reliable, relatively non-invasive, and has the potential to provide the same information as DNA extracted from more invasive sampling such as blood or tissue biopsy and proved to be suitable for restriction fragment length polymorphism, RFLP (Livia *et al*. 2006), random amplification of polymorphic DNA, RAPD and mitochondrial DNA, mtDNA (Hoolihan *et al*. 2009), microsatellite (Le Vin *et al*. 2011) studies. Brush swabbing is one of the less invasive sampling strategies, which is used in a range of other groups of organisms, e.g. amphibia (Pidancier *et al*. 2003), molluscs (Henley *et al*. 2006) and humans (Clarke *et al*. 2014). The reduced invasiveness of such sampling techniques, and reduced likelihood of impact on the welfare of the fish concerned, also makes them less likely to come under the scope of regulations on animal experimentation; swab samples can also be collected without the use of anaesthesia, required under UK regulations for tissue biopsy (Le Vin *et al*. 2011).

NGS technologies have the ability to produce millions of sequence reads to discover thousands of informative markers in a cost-effective way, applicable to quantitative, structural, functional and evolutionary studies in all organisms (Metzker 2010). NGS techniques have been used extensively in the field of molecular genetics and genomics to

address a range of biological questions since last decade. In particular, reduced representation genotyping by sequencing techniques, including restriction-site associated DNA sequencing (RADseq) in various forms (Baird *et al*. 2008; Peterson *et al*. 2012), are becoming a method of choice for many marker-based studies in non-model organisms. It is important to combine ease of DNA sampling without any harm to organisms with the cost-effectiveness of such advanced NGS technology. Quality and quantity of genomic DNA is a major concern in any NGS platform for getting high quality sequence reads. Any error within the data can yield many errors in the biological explanation of the data. Errors can arise from the sample degradation, contamination from any source (either microorganisms or cross contamination during the library preparation) or unequal representation of the DNA samples.

The use of minimally invasive DNA sampling for NGS has currently only been applied on a limited scale, e.g. in humans (Ogawa *et al*. 2013) and birds (Vo and Jedlicka 2014). As fish mucus harbours diverse microbiota and there is always chance of extracting microbial DNA, this could lead to genotyping errors due to the amplification of microbial DNA during PCR (Larsen *et al*. 2013). Some of the fish cells present may be dead, potentially reducing the integrity of the DNA and making them more prone to contamination (Lieber *et al*. 2013). It is thus important to establish if DNA derived from fish mucus is of suitable quality for NGS. It remains to be investigated if the presence of microbiota presents a significant obstacle to the routine use of skin swabs as a source of DNA for routine NGS investigation of fish genomes.

In the first part of this study, the yield and quality of DNA from skin mucus, buccal mucus (using swab brush) and fin samples of Nile tilapia, *Oreochromis niloticus* L., were assessed using two DNA extraction methods (salt precipitation and HotSHOT) and two different

types of sex-linked molecular markers (microsatellite and single nucleotide polymorphism, SNP). Bacterial DNA was also assessed by PCR. In the second part, a group of fish was monitored for 14 days following skin mucus sampling using brush and filter paper.

In the third part of the study, we investigated the potential of using DNA prepared from Nile tilapia (*Oreochromis niloticus*) skin mucus swabs for double-digest restriction-site associated DNA (ddRAD) sequencing (Peterson *et al*. 2012), by comparing genotypic output with that derived from conventional, but more invasive, sampling techniques (fin and muscle tissue biopsies).

3.3 Materials and methods

3.3.1 Sample collection

Nile tilapia were supplied by the Tropical Aquarium Facilities, Institute of Aquaculture, University of Stirling. In the first experiment (**Table 3.1**), mucus and fin samples were collected from 16 Nile tilapia individuals reared in the Tropical Aquarium, University of Stirling. Fish of approximately equal length - standard length of 12.15 ± 0.18 cm (mean ± S.D.) - and weight (33.67 ± 0.21 g) were used, and were killed humanely before sampling. A non-sterile brush (Endocervical sampler, CellPath Ltd, UK; approximately 2 cm long) was gently scraped along the lateral line or inside the mouth of the fish (10 times per individual) to collect skin (n = 16) or buccal (n = 8) mucus. Immediately after collection of skin mucus, the brushes were detached from the stem and placed in a 1.5 mL Eppendorf tube containing either 225 µL lysis solution (SSTNE buffer/1 % SDS/0.02 % proteinase K, for salt precipitation extraction of DNA; n = 8) or 100 % ethanol (fixation for later DNA extraction by salt precipitation method; n = 8), with sufficient liquid to fully immerse the brush. Buccal swab brushes were placed into lysis solution (SSTNE/SDS/proteinase K; n =

4) or alkaline lysis reagent (for HotSHOT DNA extraction; n = 4). A small portion (1.5 cm × 1.0 cm) of caudal fin tissue from each fish (n = 16) was fixed in 100 % ethanol.

In the second experiment, filter paper (2 cm × 0.5 cm) was used to collect skin mucus samples from one side (n = 10), and a brush to collect mucus from the other side (n = 10) of dorsal region of live Nile tilapia (mean standard length and weight of 16.15 ± 0.16 cm and 50.42 ± 0.50 g respectively) and were fixed into 100 % ethanol (**Table 3.1**). Following sample collection, the fish were transferred into 20 L tanks (each fish in a tank) in a recirculating system and reared in a normal condition for two weeks. Each fish was monitored two three times a day to record any behavioural changes due to sampling.

Table 3.1 Experimental design (numbers of fish; sampling and DNA extraction methods) from mucus (skin or mouth, using swab brush or filter paper) or fin tissue.

		First experiment					Second experiment	
Sampling method		Skin mucus using swab brush		Mouth mucus using swab brush		Fin tissue	Skin mucus using filter paper	Skin mucus using swab brush
DNA extraction method		Salt precipitation		Salt precipitation	HotSHOT	Salt precipitation	Salt precipitation	
		Swab out[1]	Liquid out[2]	Liquid out	Liquid out		Liquid out	Liquid out
Initial step	Lysis solution	4	4	4	4			
	Ethanol then lysis	4	4			16	10	10

[1]Swab out: Brush was removed from the tubes leaving all the solutions inside during extraction just before protein precipitation stage.
[2]Liquid out: Liquid was transferred into a new tube using a pipette during extraction just before protein precipitation stage.

In the third experiment, fin (1.5 cm × 1.0 cm), muscle (1.0 cm × 1.0 cm) and skin mucus were sampled from two fish following humane killing and preserved in 100 % ethanol for ddRADseq analysis. Filter paper was also used for skin mucus collection.

3.3.2 Genomic DNA extraction and quality check

The salt precipitation method of Aljanabi and Martinez (1997) was used with some modifications. Briefly in the first experiment, for sixteen skin swabs and four mouth swabs (see **Table 3.1**), samples were digested overnight in lysis solution and proteinase K at 55^{o}C (ethanol-fixed mucus samples were centrifuged and ethanol removed as much as possible by pipetting before digestion), then treated with RNaseA (2 mg/mL) at 37^{o}C for 1 hr. At this stage, the brush was removed in the case of eight skin swab samples (four from the group placed directly into lysis solution and four from the group that was first fixed in ethanol) while the liquid was transferred into a new 1.5 mL Eppendorf tube for the remainder skin mucus samples (four lysis solution fixed and four ethanol fixed). The liquid was transferred into the new tube at this stage in the case of four mouth swab samples. Protein was precipitated by adding 5 M NaCl. The supernatant was transferred into new tube and equal amount of absolute isopropanol was added to precipitate DNA followed by dissolving into 5 mM Tris.

Another four mouth mucus samples from the first experiment were processed with HotSHOT DNA extraction protocol (see **Table 3.1**) to extract genomic DNA (modified from Truett *et al.* 2000). Samples were heated in alkaline lysis reagent at 95^{o}C for 30 min, and then an equal volume of neutralization buffer was added to neutralize the solution. Following DNA extraction, samples were centrifuged at 21000 × G for 10 min and the supernatant (200 µL) was transferred into a new tube (the rest were thrown away) and frozen for later use.

In case of fin (c. 20 mg), muscle (c. 15 mg) and mucus (filter paper) - DNA was extracted using the salt precipitation method described above. DNA integrity was checked by Nanodrop (ND-1000) spectrophotometer and 1 % agarose gel. The concentration of double-

stranded DNA (dsDNA) was then measured more accurately; sixteen samples from skin mucus swabs, three from filter paper (in the second experiment) and the samples for ddRADseq (in the third experiment) by fluorimetry (Qubit®, Life Technologies Ltd, UK), with each sample being diluted to a standard concentration of 8 ng/µL with 5 mM Tris pH 8.0.

3.3.3 Genotyping for microsatellite marker

All the samples in the first experiment and five samples from each sampling procedure (filter paper and brush) in the second trial were analysed for a microsatellite marker (*UNH995*, Palaiokostas *et al*. 2013a) in linkage group (LG) 1, using the fluorescent labelled tailed primer genotyping assay developed by Boutin-Ganache *et al.* (2001) and described in Rajaee *et al.* (2010). In brief, 15 µL (c. 50 ng DNA) PCR reaction volumes was prepared and the thermal cyclic conditions were the initial denaturation at 95^{o}C for 1 min and 35 cycles of denaturation at 95^{o}C for 15 sec, annealing at 60^{o}C for 15 sec and extension at 72^{o}C for 30 sec (Primer sequences are in **Appendix Table C4.5**). Amplified PCR fragments were genotyped on a CEQ^{TM} 8800 capillary sequencer (Beckman Coulter®, USA) and analysed using the default parameters in the CEQ software (version 9.0).

3.3.4 Genotyping for SNP markers

All the samples in the first experiment were analysed for two SNP markers in LG1 (*Oni23063* and *Oni28137*, Palaiokostas *et al*. 2013a) and one in LG20 (*Oni3161*, Palaiokostas *et al*. 2015), while samples in the second experiment were analysed for only one SNP marker (*Oni23063*). Individuals were genotyped for SNPs using the Kompetitive Allele Specific PCR (KASP: LGC Genomics, UK) genotyping system. A 10 µL (c. 50 ng DNA) PCR reaction volume was prepared (Primer sequences are in **Appendix Table C4.4**)

and the PCR cyclic conditions were the initial denaturation at 94^{o}C for 15 min followed by 10 touchdown cycles (94^{o}C for 20 sec and touchdown 65^{o}C for 1 min, reduced by 0.8^{o}C per cycle) followed by 34 cycles of amplification (94^{o}C for 20 sec; 57^{o}C for 1 min). Following amplification, fluorescence was detected on a Techne Quantica® machine (Barloworld Scientific Ltd. UK) at ambient temperature. Allelic discrimination analysis was performed manually to determine the single nucleotide differences using the inbuilt Quansoft software (version 1.1.21).

3.3.5 DNA amplification with universal bacterial primer (16S rDNA)

In the first experiment, DNA samples from the salt precipitation method were amplified using universal bacterial primers - 16S rDNA (Hovda *et al*. 2007) to assess the bacterial DNA amplification.

A total of 5 µL PCR reaction volume was prepared with 50 ng genomic DNA (**Table 3.2**). The primer sequences are: forward primer - 5' CCTACGGGNGGCWGCAG3' and reverse primer - 5' GACTACHVGGGTATCTAATCC3'. PCR cyclic conditions were initial denaturation at 95^{o}C for 3 min, 30 cycles of denaturation at 95^{o}C for 15 sec, annealing at 55^{o}C for 15 sec and extension at 72^{o}C for 30 sec with the final extension at 72^{o}C for 10 min. A 2.5 µL of PCR products were checked on 1.5 % agarose gel electrophoresis.

Table 3.2 PCR reaction mix for universal bacterial DNA amplification.

Chemical components	**Per reaction**	**5 µL reaction**
2× MyTaq Master Mix	1×	2.5
10 uM 16S rDNA F primer	0.3 uM	0.15
10 uM 16S rDNA R primer	0.3 uM	0.15
DW	-	1.2
DNA (~50 ng/µL)		1

3.3.6 Double-digest restriction-site associated DNA (ddRAD) sequencing

3.3.6.1 Library preparation

The ddRAD library was constructed using the modified protocol described in Peterson *et al*. (2012). Briefly, each sample was replicated three times and each sample (24 ng DNA) was digested with *Sbf*I (rare cutter) and *Sph*I (frequent cutter) high fidelity restriction enzymes (NEB, UK) at 37^{o}C for 90 min, using 20 U of restriction enzyme per µg of genomic DNA in 1× CutSmart reaction buffer (NEB). Each digested DNA sample was ligated (22°C, 2.5 hrs) with Illumina-compatible individual-specific P1 (*Sbf*I compatible) and P2 adapters (*Sph*I compatible), each with a unique 5 or 7 bp barcode. Barcode sequences are provided in **Appendix Table C3.1**. Ligation was stopped by addition of 2.5 volumes PB buffer (Qiagen, UK) and the samples pooled and purified on a single column (MinElute PCR purification kit, Qiagen, UK).

Fragments were then size selected on a 1.1 % agarose gel with a portion corresponding to c. 400-700 bp being excised and gel purified (MinElute gel purification kit, Qiagen UK). This template was subjected to 11 cycles of PCR (using Q5 Hot-start High Fidelity DNA polymerase (NEB) and Illumina specific primers), the amplified library being purified twice; first by a column purification (MinElute PCR purification kit) then by a paramagnetic bead clean up (AMPure XP, Beckman Coulter, UK).

The purified amplicon library was then normalized to 10 nM stock (using EB buffer and 1 % Tween 20). The stock amplicon library (10 nM) was then again diluted to 4 nM using 2 µl of 10 nM library and 3 µl ddH_2O. The 4 nM library was denatured at room temperature for 5 min using 5 µl of freshly prepared 0.2 M NaOH. The denatured library was diluted again into 20 pM using 990 µl pre-chilled HT1 (hybridization buffer). An internal control (PhiX) was used during sequencing and PhiX was also diluted to the same loading

concentration (20 pM) as the amplicon library before loading. The final concentration of the loading library was 10.2 pM; of which 5 % was PhiX (290.7 µl of 20 pM amplicon library, 15.3 µl of 20 pM PhiX library and 294.0 µl of HT1 buffer). The final library was heat denatured at 98^{o}C for 2 min followed by cooling in an ice bath for 5 min before loading on an Illumina MiSeq cartridge. A flow cell was loaded onto the MiSeq machine first according to the manufacturer's instruction. The denatured library was loaded on an Illumina MiSeq cartridge and was sequenced as part of a shared run on Illumina MiSeq platform using v2 sequencing chemistry in a single lane for 300 cycles (161 bases paired-end reads).

3.3.6.2 Generation of RAD loci and further analysis

The quality of the generated sequence reads was initially assessed using FASTQC software (Version 0.11.2). Raw reads were processed and demultiplexed using the default parameters of the process_radtags component in STACKS (version 1.27), a software pipeline designed for RAD-based analysis (Catchen *et al*. 2013). Filters removed low quality reads (process_radtags – s parameter set to 20), reads missing restriction enzyme cut sites together with those reads with ambiguous or unpaired barcodes. Retained reads were sorted into RAD loci using both reference–based and *de novo* assembly approaches (employing ref_map.pl and denovo_map.pl components respectively). For the reference-based analysis, sequence alignment/map (SAM) files were created using the Bowtie aligner (Langmead 2010) utilising the published tilapia genome (Broad Institute of MIT and Harvard genome assembly Orenil1.1; Brawand *et al*. 2014). The main STACKS parameters used for *de novo* analysis were as follows: minimum stack depth (m) = 6; mismatches allowed between stacks (M) = 2 and mismatches allowed between catalog loci (n) = 1. For reference-based analysis the parameters were m = 10 and n = 1.

The identified RAD loci from *de novo* assembly analysis were blasted against published Nile tilapia genomes, bacterial genomes and against NCBI nt database using Blastn analysis (Blastn 2.2.28+, Altschul *et al*. 1990) to identify the sequence homology. All the RAD loci from the reference–based and *de novo* assembly approaches were used to identify the shared loci (http://bioinfogp.cnb.csic.es/tools/venny/) and the results were visualised as Venn diagrams.

Data access

The raw sequence data for this study have been submitted to the EBI's European Nucleotide Archive (ENA), Sequence Read Archive (SRA) - study accession number PRJEB13792.

3.4 Results

3.4.1 First experiment

In the first experiment, DNA quality and quantity from swab brush was compared with the DNA from fin sample using two different DNA extraction methods and two types of molecular markers.

3.4.1.1 Genomic DNA quantity and quality

Genomic DNA was successfully extracted from body mucus, buccal mucus and fin samples of all tested individuals irrespective of DNA extraction method. Total DNA yield was in the ranges 13 to 39 µg and 7 to 23 µg based on the Nanodrop and Qubit quantification method respectively. OD260/OD280 ratio was 1.8 to 1.9 and OD260/OD230 was 2.0 to 2.2, which indicate good quality DNA (**Table 3.3**). DNA quality was also checked on 1 % agarose gels with no visible sign of RNA contamination or DNA degradation (**Figure 3.1**). Based on

Nanodrop results, fixation in ethanol and removing the swab brush from the digestion significantly reduced the DNA yield from the skin mucus swab ($P < 0.05$; two-way analysis of variance (ANOVA) using Minitab 16 statistical software, **Table 3.3**). There was no significant effect of using different storage and separation methods on DNA purity. On the other hand, separation methods (swab out or liquid out) during extraction of DNA from body mucus had significant statistical differences ($P < 0.05$) on quantity of DNA but the storage methods (either lysis solution or ethanol) and their interaction had no effect on DNA concentrations based on Qubit (**Table 3.3**).

Table 3.3 Total DNA yield (µg) by using Nanodrop spectrophotometer and Qubit fluorimetry from skin mucus swab using different storage and separation methods in the first experiment.

Storage	**NanoDrop Separation (mean ± SD)**		**Qubit Separation (mean ± SD)**	
	Swab out	Liquid out	Swab out	Liquid out
Lysis Solution	24.80 ± 3.48	39.24 ± 6.21	14.03 ± 3.20	23.81 ± 6.05
Ethanol	13.27 ± 5.75	19.19 ± 6.66	7.73 ± 2.40	16.93 ± 8.76

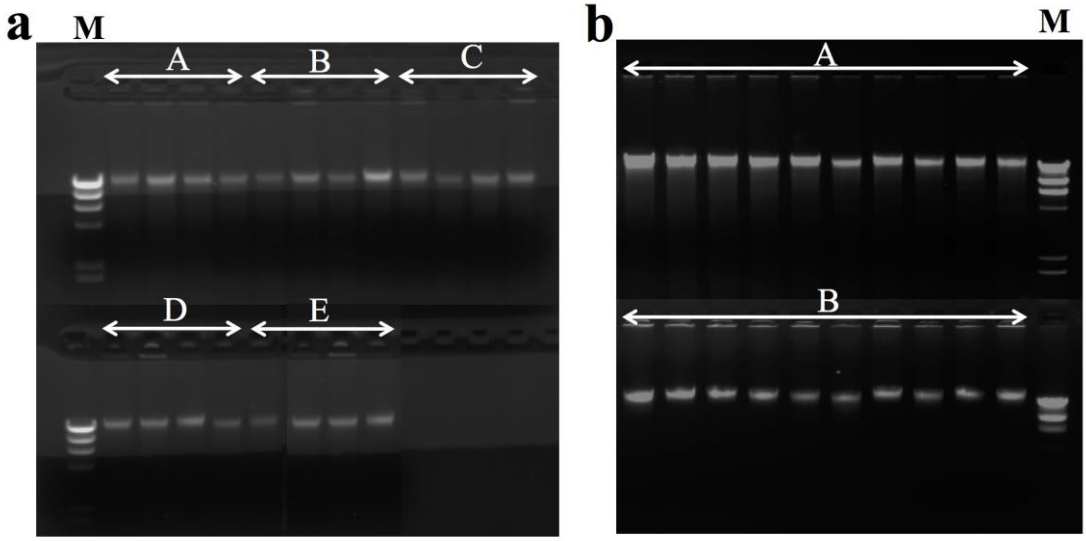


Figure 3.1 Integrity of genomic DNA (c. 50 ng) extracted using salt precipitation method on 1 % agarose gel. a) DNA samples from different storage and separation methods used in the first experiment. A - lysis solution swab out, B - ethanol swab out, C - lysis solution liquid out, D - ethanol liquid out, E - mouth swab out. b) DNA samples from filter paper (A) and brush swab (B) used in the second experiment. M - molecular weight ladder (λ DNA-Hind III).

3.4.1.2 Microsatellite and SNP markers analysis

The microsatellite marker (*UNH995*) genotyping results from mucus and fin samples from the same fish using salt precipitation extraction showed 100 % similarities on gel electrophoresis and genotyping outputs (**Figure 3.2**, **3.4**; **Table 3.5**). However, large allele dropout was observed using DNA samples extracted following the HotSHOT method (reduced intensity of the larger 236 bp allele in heterozygotes, leading to the band being absent when scored by eye on agarose gels and very low peak heights in the corresponding sequencer output (**Figure 3.3**, **3.4a**).

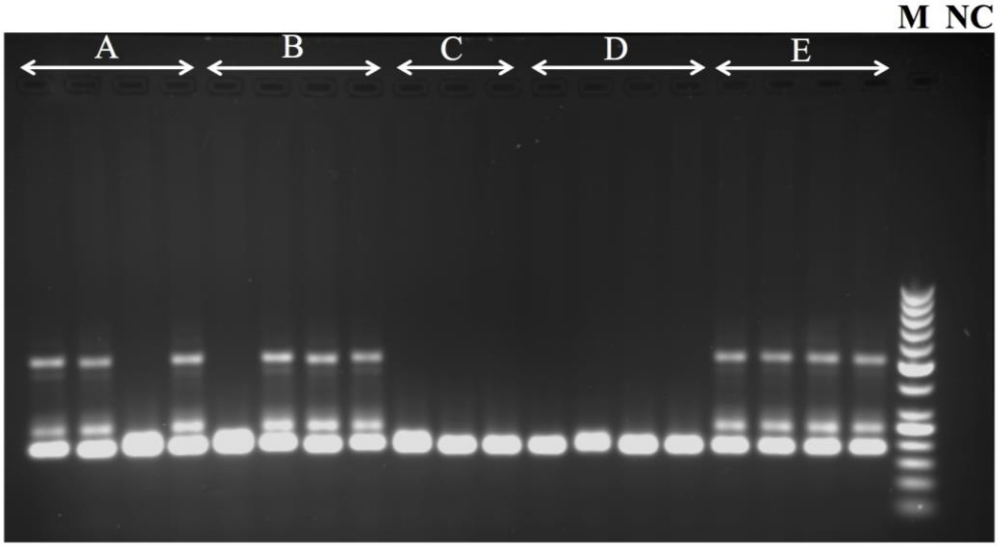


Figure 3.2 Microsatellite locus (*UNH995*) amplified PCR products on 1.5 % agarose gel. A - lysis solution swab out (1-4), B - lysis solution liquid out (1-4), C – fin (2-4), D - ethanol swab out (1-4), E - ethanol liquid out (1-4). M - molecular weight ladder (50 bp), NC – negative control.

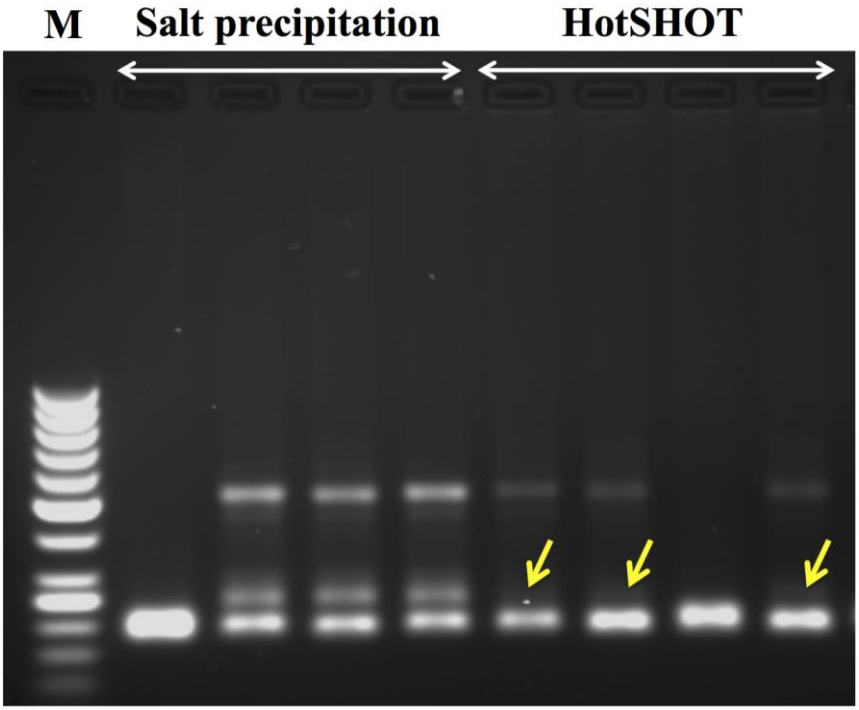


Figure 3.3 1.5 % agarose gel electrophoresis of PCR products for microsatellite locus (*UNH995*) to evaluate the efficiency of two different DNA extraction methods (salt precipitation and HotSHOT) from buccal swabbing. Dropout of the larger allele in HotSHOT method (arrows). M - molecular weight ladder (50 bp).

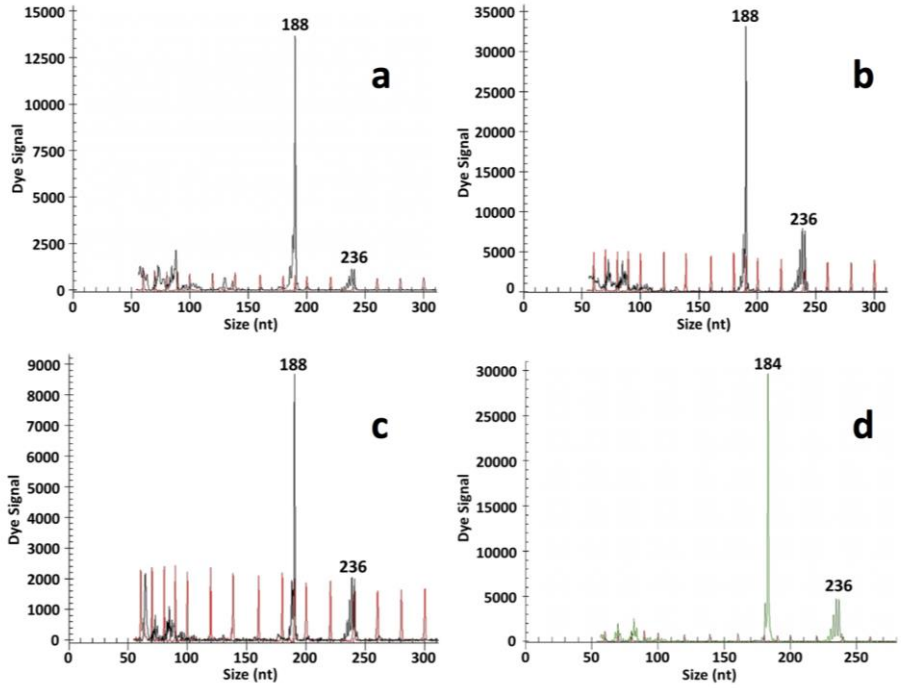


Figure 3.4 Comparison of microsatellite electropherograms of genomic DNA extracted by using (a) brush from mouth (HotSHOT), (b) brush from skin, (c) fin and (d) filter paper from skin amplified at microsatellite locus *UNH995*. Dropout of the larger allele (236 bp) on the agarose gel (HotSHOT method, yellow arrow in **Figure 3.3**) was evident on sequencer (a).

Mucus samples also gave consistent results with the fin samples for SNP markers irrespective of the DNA extraction process (either salt precipitation or HotSHOT, **Figure 3.5**).

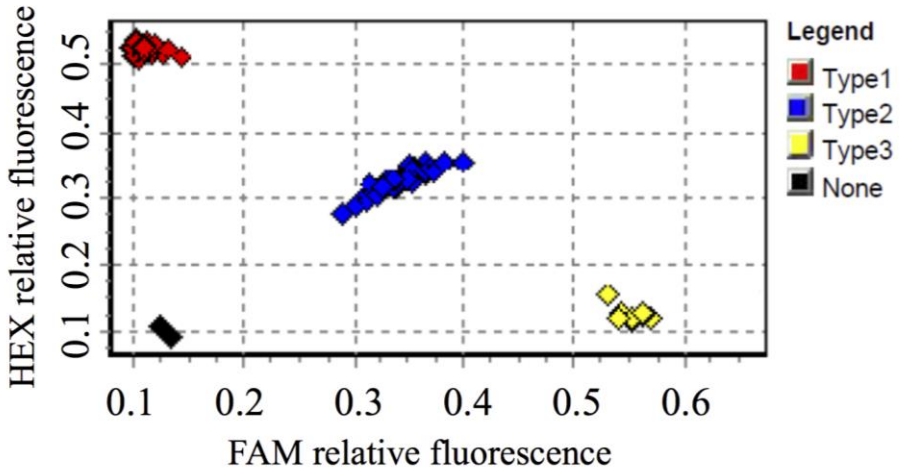


Figure 3.5 Graphical representation of the KASP assay results: body mucus, buccal mucus and fin samples from the same fish using filter paper and brush swab including other few fin samples with positive and negative controls were analysed using three SNP markers. Type1 - homozygote for HEX allele, Type3 - homozygote for FAM allele Type2 - heterozygotes for both alleles and None- ddH_2O.

3.4.1.3 Bacterial DNA analysis

Higher amounts of bacterial DNA were amplified using 16S rDNA from mucus samples than from the fin samples (**Figure 3.6**), although it should be stressed that this assay was not accurately quantitative. More bacterial DNA was amplified from buccal mucus than from skin mucus.

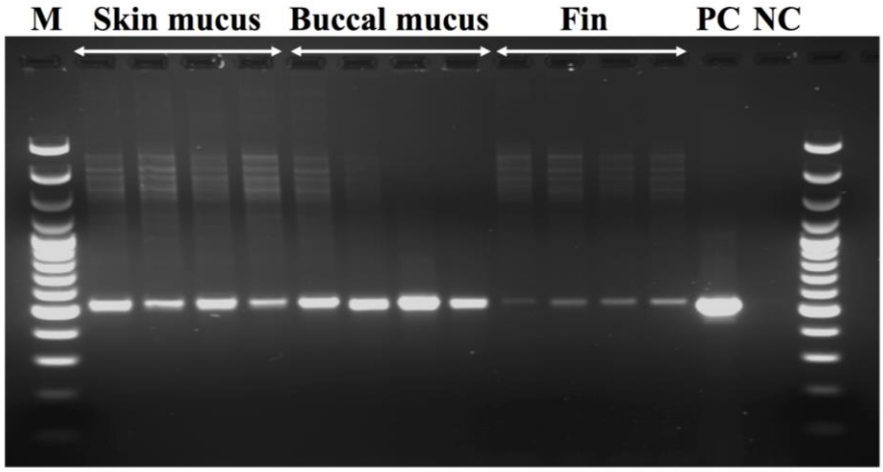


Figure 3.6 Universal bacterial primer (16S rDNA) was amplified from DNA extracted from mucus (skin and buccal) and fin samples. PCR products was run on 1.5 % agarose gel including positive (PC) and negative (NC) controls. M - molecular weight marker (100 bp).

3.4.2 Second experiment

In the second experiment, DNA quantity and quality from swab brush was compared with the filter paper.

Total DNA yield from filter paper and brush were 4 and 20 µg respectively based on Nanodrop (0.4 to 0.7 µg from filter paper based on Qubit). The OD260/OD280 ratio was 1.8 to 1.9; and OD260/OD230 was 2.16 to 2.28 for the samples collected using filter paper and brush (**Table 3.4**). There were significant differences ($P < 0.05$) between filter paper and brush in the quantity and quality (absorbance ratios) of the DNA extracted. Although filter paper yielded less DNA than swab brush, DNA derived from filter paper showed good quality DNA in both absorbance ratio (Nanodrop, **Table 3.4**) and high molecular weight on the gel electrophoresis (**Figure 3.1b**).

Table 3.4 Total DNA yield (µg) and absorbance ratio by using Nanodrop spectrophotometer from the filter paper and brush swab in the second experiment.

	Total DNA yield (µg)	**OD260/OD280 (mean ± SD)**	**OD260/OD230 (mean ± SD)**
Filter paper	4.18 ± 2.14	1.86 ± 0.02	2.16 ± 0.11
Brush swab	20.52 ± 9.85	1.91 ± 0.02	2.28 ± 0.12

The same genotypes were found for microsatellite (*UNH995*) and SNP markers (*Oni23063*) from filter paper and brush swab (**Table 3.5**). There were no evident of mortality or damage on the skin of the fish or any behavioural changes during the 14 days of rearing period.

Table 3.5 Genotypic results for microsatellite (*UNH995*) and SNP markers (*Oni23063*, *Oni28137* and *Oni3161*) for each sample based on the separation, storage and DNA extraction process in the first and second experiment.

	Microsatellite	**SNP**		
Sample ID	***UNH995***	***Oni23063***	***Oni28137***	***Oni3161***
1F/ES1	184/184	G/G	T/T	C/T
2F/ES2	184/188	G/G	T/T	C/C
3F/ES3	184/184	G/G	T/T	C/T
4F/ES4	184/184	G/G	T/T	C/T
5F/EL1	184/236	A/G	G/T	C/C
6F/EL2	184/236	A/G	G/T	C/C
7F/EL3	184/236	A/G	G/T	C/T
8F/EL4	184/236	A/G	G/T	C/T
9F/FS1/MHS1	188/236	A/G	G/T	C/T
10F/FS2/MHS2	184/236	A/G	G/T	C/C
11F/FS3/MHS3	184/188	G/G	T/T	C/C
12F/FS4/MHS4	184/236	A/G	G/T	C/T
13F/FL1/MS1	184/188	G/G	T/T	C/T
14F/FL2/MS2	188/236	A/G	G/T	C/C
15F/FL3/MS3	184/236	A/G	G/T	C/C
16F/FL4/MS4	188/236	A/G	G/T	C/T
1FP/1B	184/190	G/G		
2FP/2B	184/236	A/G		
3FP/3B	184/236	A/G		
4FP/4B	184/184	A/A		
5FP/5B	184/184	A/A		

3.4.3 Third experiment

The third experiment was conducted to investigate the potential use of mucus-derived DNA for ddRADseq and to determine if the level of bacterial DNA in the mucus samples (found in the first experiment) has any negative impact on the sequencing outputs.

Total DNA yield from muscle, mucus and fin were c. 3.0, 9.5, 18.0 µg respectively (based on Nanodrop) and the purity was reasonable (OD260/OD280 ranged between 1.8 and 1.9).

3.4.3.1 RAD sequencing

In total 6 samples were analysed and samples were barcoded, pooled and sequenced on the Illumina MiSeq sequencer. A total of 4,159,392 reads (2,079,696 paired-end reads) were assigned to the six samples following sequencing with 91.8 % of these reads (total 3,819,058) being retained after filtering out of low quality reads and reads lacking barcodes or restriction enzyme cut sites for further analysis (EBI, SRA study accession number PRJEB13792). The number of reads per sample was relatively consistent ranging between 511 K - 806 K (**Table 3.6**).

The reference-based STACKS analysis identified c. 9 K unique RAD loci (mean read depth c. 72×) with very similar numbers of loci being identified from fin, muscle and mucus DNA samples in each fish. More than 1 K RAD loci were polymorphic and the numbers were very similar among the sample types in both fish, and approximately 12-13 % of the total RAD loci were identified as bi-allelic (**Table 3.6**).

De novo based analysis of RAD loci was carried out at a lower read-depth threshold (m= 6; *cf*. m=10 for reference-based analysis) to increase the likelihood of detecting low levels of possible contaminating microbiota. Similar numbers of RAD loci (c. 14 K, mean depth c. 44×) were found in each sample for both fish as was the case in the reference-based

approaches. Very similar number of polymorphic RAD loci (> 3 K) and bi-allelic RAD loci (> 19 %) were retrieved among the samples (**Table 3.6**).

3.4.3.2 Homology searches

It was found that the vast majority of the RAD loci were common (qualitatively) to all three DNA samples (mucus, muscle and fin). Approximately > 7.5 K RAD loci and > 13 K RAD loci were shared in each fish from the reference-based (total c. 9 K RAD loci) and the *de novo* approaches (total c. 15 K RAD loci) respectively (**Figure 3.7**).

Blastn comparisons (e-value $\leq 10^{-20}$) showed that > 98 % of RAD loci were present in the published Nile tilapia genome (**Table 3.7**). The remaining sequences were then blasted against all bacterial genomes available at NCBI (April, 2015) where only 0.07 % of the total RAD loci gave positive hits (e-value $\leq 10^{-20}$), identifying a number of different Actinobacteria. These were found in just one of the muscle samples. No bacterial sequences were identified in either skin mucus or fin-derived DNA samples. Blasting the remaining loci against the NCBI nt database identified additional teleost fish sequences (c. 0.42 % of total RAD loci; **Table 3.7**). Approximately 1.25 to 1.49 % of the RAD loci failed to show homology to any of the databases searched; these largely comprising repetitive type sequences.

Table 3.6 Details of the number of reads after filtering followed by reconstructed number of RAD loci (using reference-based and *de novo* genome assembly), polymorphic loci and bi-allelic loci including percentages from fin, muscle and mucus for two fish.

	Reference-based analysis						***De novo* analysis**					
	Fish 1			**Fish 2**			**Fish 1**			**Fish 2**		
	Fin	**Muscle**	**Mucus**	**Fin**	**Muscle**	**Mucus**	**Fin**	**Muscle**	**Mucus**	**Fin**	**Muscle**	**Mucus**
Filtered reads	600,780	805,584	555,466	710,454	510,838	635,936	600,780	805,584	555,466	710,454	510,838	635,936
RAD loci	8,163	8,984	8,341	8,647	7,891	8,519	14,135	15,261	14,354	14,723	13,729	14,505
Polymorphic loci	1,181 (14.47 %)	1,296 (14.43 %)	1,216 (14.57 %)	1,148 (13.28 %)	1,038 (13.15 %)	1,133 (13.30 %)	3,230 (22.85 %)	3,486 (22.84 %)	3,318 (23.12 %)	3,426 (23.27 %)	3,198 (23.29 %)	3,389 (23.36 %)
Bi-allelic loci	1,114 (13.65 %)	1,220 (13.58 %)	1,143 (13.70 %)	1,082 (12.51 %)	976 (12.37 %)	1,063 (12.48 %)	2,754 (19.48 %)	2,961 (19.40 %)	2,821 (19.65 %)	2,919 (19.83 %)	2,711 (19.75 %)	2,877 (19.83 %)

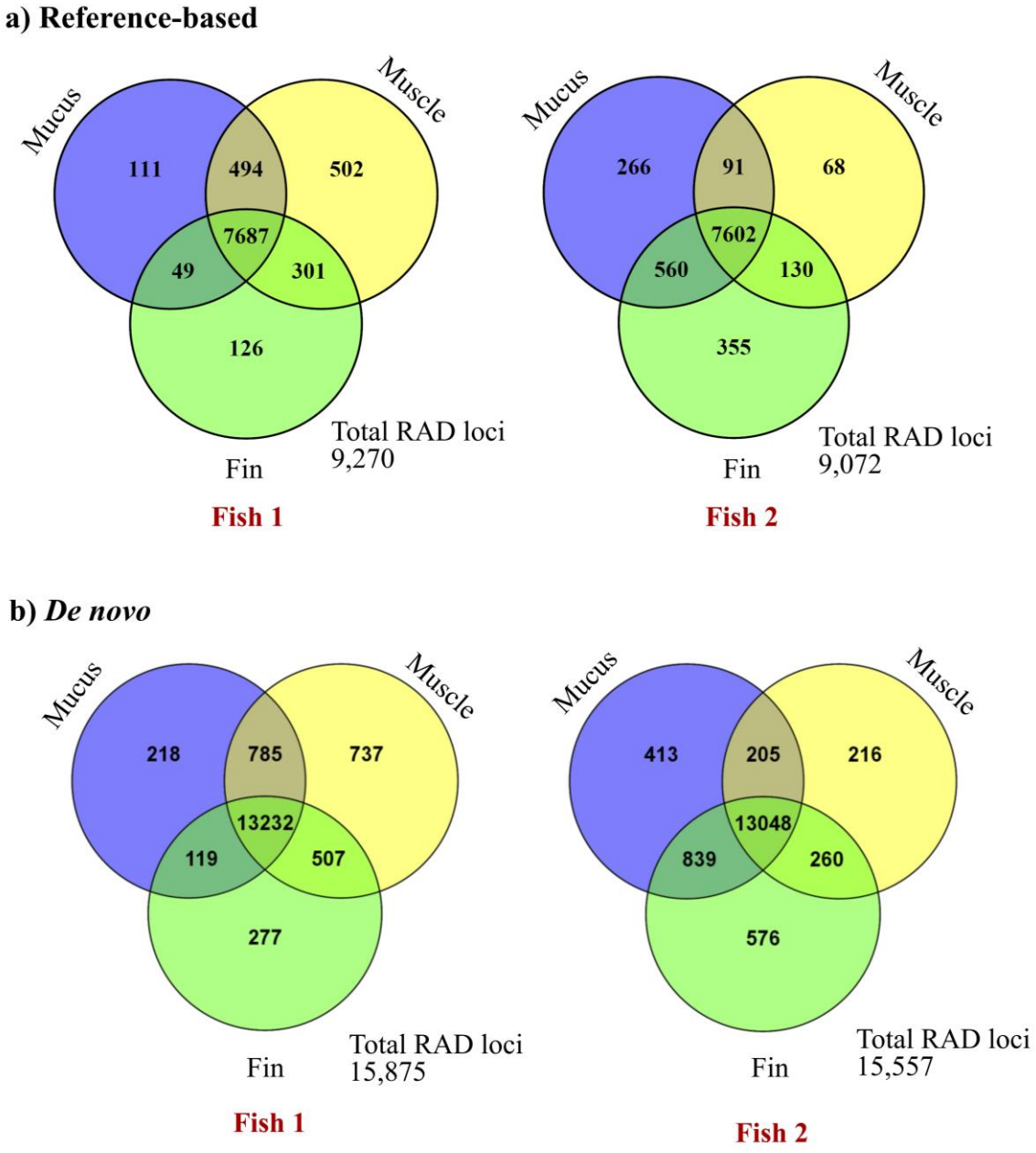


Figure 3.7 Venn diagrams depicting the partitioning of numbers of RAD loci identified from mucus, muscle and fin tissues from each fish; (a) reference-based genome assembly, (b) *de novo* genome assembly.

Table 3.7 Homology searches of RAD loci generated from *de novo* genome assembly against NCBI databases using Blastn.

	Fish 1			**Fish 2**		
	Fin	**Muscle**	**Mucus**	**Fin**	**Muscle**	**Mucus**
RAD loci	14,135	15,261	14,354	14,723	13,729	14,505
Database matches						
Nile Tilapia genome	13,876 **(98.2 %)**	14,966 **(98.1 %)**	14,079 **(98.1 %)**	14,471 **(98.3 %)**	13,458 **(98.0 %)**	14,262 **(98.3 %)**
Bacterial genomes	0	0	0	0	10 (0.07 %)	0
nt database	54 (0.38 %)	67 (0.44 %)	67 (0.47 %)	63 (0.43 %)	62 (0.45 %)	62 (0.43 %)
No hit	205 (1.45 %)	228 (1.49 %)	208 (1.45 %)	189 (1.28 %)	199 (1.45 %)	181 (1.25 %)

3.5 Discussion

Skin mucus was found to be suitable source of DNA for standard genotyping and ddRADseq techniques in the Nile tilapia, and that sampling can be carried out using swab brush or filter paper without causing any damage to the fish. No physical damage, lesion or behavioural changes were found due to the sampling process.

At least the required quantity of DNA, of adequate quality, was extracted from different sources of storage and extraction processes in both brush swab and filter paper. Ethanol storage and removal of the swab brush during the extraction process yielded significantly less DNA compared to the liquid removal or lysis solution storage. It has been noticed that significant amount of solutions were lost while removing the brush (solutions attached to the brush) which could be the reason to yield less DNA from "swab brush out" samples

during the extraction process. The total yield from either method was adequate for PCR-based molecular DNA analysis.

In general the amount of tissue/DNA required for genetic/genomic analysis has declined significantly over the last few decades. For example in the 1970s and 1980s, the Southern transfer was one of the most common techniques, and might require analysis of DNA from an individual using separate DNA aliquots of 5-10 μg each, digested with different restriction enzymes, so we might have looked to extract at least 50 μg per individual (Melchers *et al*. 1989). Now a day about 0.05 to 0.10 μg DNA has been found to suitable for PCR-based microsatellite marker analysis (Rajaee *et al*. 2010).

A significant difference was observed between filter paper and swab brush sampling in terms of the total DNA yield and DNA quality. Total DNA yield was lower from filter paper (which is adequate for molecular analysis) than from the swab brush but the quality of the DNA was better from filter paper-derived DNA.

Filter paper is cheaper and in some ways more convenient to use (e.g. it is easier to extract DNA from a proportion of the fixed sample using filter paper than a swab brush) and the salt precipitation method extracted good quality DNA (no evident of RNA contamination) that can be used for more sophisticated and sensitive NGS techniques.

High molecular weight DNA was extracted from all the mucus samples irrespective of sampling and DNA extraction process with no indication of DNA degradation. A slight degradation of DNA was noticed in mucus-derived DNA compared to the fin clips in one earlier study (Le Vin *et al*. 2011).

Mucus samples gave consistent results with fin samples for microsatellite (*UNH995*) and SNP (*Oni23063*, *Oni28137* and *Oni3161*) markers, irrespective of sample collection and

extraction method (brush removal/ethanol fixation). The consistent results also suggested that there was no cross-contamination of samples during mucus sampling (fish kept in the same tank during sample collection) and DNA analysis. These results support the findings of Mirimin *et al*. (2011), who also found 100 % similar genotyping output from mucus with that of the fin clips. Some other studies have been found some genotyping mismatches and allelic dropout in microsatellite markers analysis from buccal or body mucus compared to fin clips in small and large fish (Livia *et al*. 2006; Reid *et al*. 2012; Le Vin *et al*. 2011). Contamination from other individuals at high density environment is thought to give erroneous genotyping results. Le Vin *et al*. (2011) conducted an experiment using mucus samples collected from low density (2 fish/150 L tank) and high density (3-19 fish/50 L tank) group of fishes and compared with fin clips. They mentioned no evidence of cross contamination between individuals (similar genotyping output) kept at high density.

Large allelic dropout was observed for the microsatellite marker on agarose gel electrophoresis in samples using the HotSHOT DNA extraction protocol, which in turn was evident in the sequencer with very low peak heights. Meissner *et al.* (2013) also confirmed that allelic dropout was evident with the HotSHOT method, using DNA extracted from gill tissue and stored for 22 days (samples were stored for 12 days in the present study).

More bacterial DNA amplification was observed in the mucus-derived DNA than the DNA derived from fin. The presence of bacterial DNA did not hinder the genomic DNA amplification using microsatellite and SNP specific primers under normal conditions for such markers. Hoolihan *et al*. (2009) reported that mucus-derived DNA gave a slight different results in RAPD amplification profile from the muscle-derived DNA and they postulated the presence of bacterial or other micro-algal DNA in the mucus-derived DNA or

other artefact variations common for RAPD technique (Ellsworth *et al*. 1993) could be the reason for the different amplification results.

Skin mucus swabs have been used as a source for fish DNA in many targeted single locus PCR-based studies (Hoolihan *et al*. 2009; Le Vin *et al*. 2011), but their suitability for assaying genome-wide variability through RADseq and related next generation sequencing methodologies has not been examined. Fish mucus has lots of microorganisms and higher bacterial DNA amplification was demonstrated (in the first experiment) in tilapia DNA extracted from skin mucus swabs compared to the DNA extracted from fin samples, so there was potential for extraneous DNA from microbiota in swab samples to affect NGS analysis. The third experiment was thus designed to compare DNA derived from mucus, fin and muscle samples from the same fish in ddRADseq analysis.

The sequencing generated raw reads that were very similar among the three tissue types in both fish, which means the quality and quantity of genomic DNA derived from mucus was similar to fin and muscle-derived DNA. Similar results were found from the reference-based and *de novo* assembly approaches in three different sample types for both fish in terms of the number of generated RAD loci, number of polymorphic loci and the numbers of bi-allelic RAD loci. Approximately 12-13 % and > 19 % RAD loci were bi-allelic based on reference-based and *de novo* assembly approaches respectively. Palaiokostas *et al*. (2015) also generated 12.4 % bi-allelic RAD loci (reference-based genome assembly) from the fin-derived DNA samples using ddRADseq from the same stock of Nile tilapia.

Quantitatively the results were similar in the three tissue types and > 83 % of the RAD loci retrieved were shared among mucus, muscle and fin-derived DNA samples. More than 98 % of total RAD loci (both reference-based and *de novo* assembly) were present in the published Nile tilapia genome.

In *de novo* assembly approach, a lower read depth coverage (m = 6) compared to the reference-based approach (m = 10) was used to increase the chances of detecting the presence of bacterial sequences in Blast searches. Bacterial sequences were only detected in muscle-derived DNA (0.07 %), showing that although bacterial DNA appeared to be present in higher concentration in DNA derived from mucus than that derived from fin tissue, no detectable amount of this appeared in NGS-generated RAD loci.

DNA extracted from skin mucus swabs was used successfully in ddRAD sequencing, with no discernible differences in ddRAD data quality or composition generated from this source being evident, compared to DNA from other tissues. Minimal contamination from bacterial or other sources was apparent and not detected at all in skin mucus samples. This small NGS-mediated genomic analysis, based on minimally invasive DNA sampling, should encourage fish geneticists working in a range of disciplines (molecular ecology, conservation and aquaculture) to use such sampling for NGS techniques.

3.6 Conclusions

Skin mucus is a suitable source of fish genomic DNA and a viable alternative to fin or other more invasive samples, for at least this type of genetic and genomic analysis, such minimally invasive sampling should have wide applicability in research and management of fish populations.

Acknowledgements

The authors acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance (CSC reference BDCA-2013-5) and Keith Ranson, Brian Craig, Kerry

Bartie and Sarah-Louise Counter for their technical support in the Tropical Aquarium Facilities and in the Laboratory at the Institute of Aquaculture, University of Stirling. The work was also partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions.

Chapter 4

Sex determination in GIFT

Analysis of sex determination in genetically improved farmed tilapia

Taslima K.[1,2], Taggart J.B.[1], Wehner S.[1], de Verdal H.[3,4], Benzie J.[3], Bekaert M.[1], McAndrew B.J.[1], Penman D.J.[1]

[1]Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2]Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh
[3]WorldFish Centre, Jalan Batu Maung, 11960 Bayan Lepas, Penang, Malaysia
[4]Cirad, UMR ISEM, 34398 Montpellier Cedex 5, France

Status: To be submitted

Contributions:

Experimental design, sample preparation for BSA, genomic DNA extraction, construction of the ddRAD libraries and sequencing (under the supervision of John B. Taggart), DNA marker genotyping, sample preparation for Sanger sequencing, association analysis with different markers and general statistics were conducted by the author of this thesis. Data compiling and the manuscript were first drafted by the author of this thesis who was also entirely involved in the subsequent corrections of the manuscript. The other co-authors contributed to the experimental design, breeding of fish, gonad squash for sexing and fin sample collection, analysis of generated sequences from MiSeq and editing the manuscript.

4.1 Abstract

The genetically improved farmed tilapia (GIFT) stock, developed by the WorldFish Center, is the single most important resource for tilapia aquaculture. It was founded from multiple wild and domesticated populations of Nile tilapia (*Oreochromis niloticus*), with some of the domesticated stocks likely to have been introgressed with *O. mossambicus*. Different major sex-determining loci have been detected in different tilapia stocks, the major influence in GIFT is currently unknown. A bulk segregant analysis (BSA) version of double-digest restriction-site associated DNA sequencing (BSA-ddRADseq) was used to detect and map sex-linked single nucleotide polymorphisms (SNP) in 19 GIFT stock families, with two Stirling stock families (previously determined to have balanced sex-ratios controlled by an XX/XY locus in linkage group 1) as controls. DNA was extracted from fin tissues of individual sires and dams and from pooled fin samples of progeny, segregated according to phenotypic sex. ddRAD sequencing was performed on the parental and sex-specific pooled progeny, identifying, scoring and analysing segregation patterns in approximately 1,500 SNPs per family across the genome. Phenotypic sex in Stirling (control) families showed strong association with LG1 (as expected), whereas only SNPs located in LG23 showed clear association with sex in the majority of the GIFT families. No other genomic regions linked to sex determination were apparent from the analysis. In order to validate this result, progeny from six GIFT families (three showing the strongest association with LG23 and three showing the weakest or no association) and 50 broodstock were individually genotyped for a series of LG23-specific DNA markers: five SNPs showing the highest association with sex in BSA-ddRADseq analysis, the LG23 sex-associated microsatellite *UNH898* and *ARO172*; and the recently isolated *Amh*Δ*y* (containing two insertions and four deletions relative to *Amh*) and *Amhy* (one deletion and a single SNP) marker (both have been shown to be Y-specific in some stocks of this species). All of these markers showed

high association with sex in all six families. This is the first genomic analysis of sex determination in the GIFT stock, with sex appearing to be predominantly determined by a locus in LG23. BSA-ddRADseq provided an efficient and cost-effective means to establish the position of the major sex-determining region in GIFT tilapia. The sex-linked markers identified will be useful for potential marker-assisted selection (MAS) to control sex-ratio in GIFT tilapia, to control unwanted reproduction during growout.

Keywords: Genetically improved farmed tilapia (GIFT), sex determination, BSA, ddRADseq, MAS, sex-ratio control.

4.2 Introduction

Sex determination and differentiation are considered to be a fundamental step in the life of an organism. In animals with male (XX/XY) and female (WZ/ZZ) heterogametic sex determining systems, a master sex-determining gene in the regulatory cascade triggers the development of either sex (male/female). Nearly all mammals have a male heterogametic sex-determining system (XX/XY) and Y-linked *Sry* (sex-determining region of the Y chromosome), the first isolated vertebrate master sex-determining gene, regulates sex determination in mammals (Sinclair *et al.* 1990; Koopman *et al.* 1991). On the other hand, female heterogamety (WZ/ZZ) has been observed in all birds and some snakes where Z-linked *Dmrt1* (doublesex and mab-3 related transcription factor1) triggers male sex development in a double dosage mechanism (Ezaz *et al.* 2006; Siegfried 2010). Simple (male or female heterogametic) to complex (polygenic) chromosomal sex determination, environmental sex determination and sometimes interaction between genes and environmental factors have been observed in fish, lizards, turtles and amphibians (Ezaz *et al.* 2006).

Fishes are an immensely diverse group of species and exhibit an exceptional variety of reproductive strategies, which are categorized into gonochorism, synchronous/sequential hermaphroditism or unisexual reproduction. Sex determination in fish does not appear to be strongly conserved from one group to another and varies among closely related species, even showing intraspecific variation (Desjardins and Fernald 2009; Sandra and Norma 2010). For example three different genes responsible for sex determination have been identified in three different fish species in one genus: *Dmy/dmrt1by* in *Oryzias latipes* (Matsuda *et al.* 2002; Nanda *et al.* 2002), *Sox3y* (*Sry* related gene) in *O. dancena* (Takehana *et al.* 2014) and *Gsdfy* (gonadal soma-derived growth factor on the Y chromosome) in *O. luzonensis* (Myosho *et al.* 2012). Different components of the transforming growth factor beta (TGF-β) were found as strong candidates for master sex-determining genes in different fish species - *Amhy* (duplicate of anti-Müllerian hormone on the Y chromosome) in *Odontesthes hatcheri* (Hattori *et al.* 2012), *AmhrII* (a SNP in *Amh* receptor type II) in *Takifugu rubripes* (Kamiya *et al.* 2012) and *Amhy* (duplicate of *Amh* gene on the Y chromosome) in *O. niloticus* (Eshel *et al.* 2014; Li *et al.* 2015). A sexually dimorphic immune-related gene only present on the Y chromosome (*Sdy*) evolved as the master sex-determining gene in the rainbow trout, *Oncorhynchus mykiss* (Yano *et al*. 2012) and this male-specific gene has found to be conserved across the salmonids (Yano *et al*. 2013).

Tilapia shows great species diversity, with more than 70 freshwater and a few brackish water species. The Nile tilapia (*O. niloticus*) plays a significant global role in commercial aquaculture production. Nile tilapia originates from Africa and since the introduction of Nile tilapia from Africa to different Asian countries, the genetic quality of the original stocks often deteriorated because of genetic founder and bottleneck effects followed by the probable inbreeding depression, owing to the import of limited numbers of fish from Africa and low effective population sizes (Bentsen *et al*. 1998). In addition to these, the purity of

Nile tilapia aquaculture stocks (e.g. in Asian countries) deteriorated due to introgression from the less desirable Mozambique tilapia, imported for aquaculture and feral in many countries (Macaranas *et al*. 1986; Romana-Eguia *et al*. 2005). To improve the genetic quality of farmed stocks of this species, and more generally to demonstrate the potential for genetic improvement in warm water aquaculture, the genetically improved farmed tilapia (GIFT) was developed by the WorldFish Centre through selective breeding. The GIFT base population was developed from multiple wild (African) and domesticated (Asian) populations of Nile tilapia (*O. niloticus*). This has made a significant contribution to world tilapia aquaculture production, and its success has led to many other selective breeding programmes.

Both male (XX/XY) and female (WZ/ZZ) heterogametic sex-determining systems and environmental influences on sex are evident in different tilapia species. The variety of sex-determining systems in tilapia and the demand for single sex (monosex male) culture systems (not only for growth but also to avoid unwanted reproduction) have encouraged researchers to elucidate sex determination in tilapia. Different sex-determining loci have been mapped in different chromosomes (linkage groups, LG) in tilapia. From microsatellite marker-based studies, it has been found that LG1 and LG3 were linked to the phenotypic sex in blue tilapia, which possesses primarily female heterogametic sex determination (Lee *et al*. 2004), whereas LG1 was found to be linked to phenotypic sex in Mozambique tilapia originated from South Africa (Liu *et al*. 2013). A very recent study has identified a gene, *OsZfand3* (Zinc finger AN1-type domain 3) in LG1 in hybrid tilapia (crosses from Mozambique and red tilapia), which is thought to be tightly linked to the sex-determining locus (Ma *et al*. 2016).

Nile tilapia (*O. niloticus*) possesses male heterogamety (XX/XY) which is controlled by major genetic factors and sometimes may interact with minor genetic or environmental factors (Penman and Piferrer 2008). Two different XX/XY sex-determining loci (in LG1 and LG23) have been mapped in different stocks of Nile tilapia. A major sex-determining locus was mapped to LG1 in the Stirling strain of Nile tilapia, originally derived from Lake Manzala in Egypt, using BSA-mediated microsatellite marker analysis (Lee *et al*. 2003) and restriction-site associated DNA sequencing (RADseq) (Palaiokostas *et al*. 2013a). Thermosensitivity has also been observed in Nile tilapia, associated with a locus in LG20 (Palaiokostas *et al*. 2015). Family-specific QTLs have also been identified in LG1, LG3 and LG23 for temperature-dependent sex in a Nile tilapia population derived from the Stirling strain (Lühmann *et al*. 2012; Wessels *et al*. 2014).

In a stock in Israel, derived from the Swansea stock of Nile tilapia (also derived from Stirling), another XX/XY sex-determining locus, in LG23, has been found using simple sequence repeats (SSR) and sex-specific markers analysis (Eshel *et al*. 2011, 2012). Recently a duplicated copy of *Amh* gene, *Amhy* (associated with male sex determination) was identified as a candidate sex determiner in this stock (Eshel *et al*. 2014). Li *et al*. (2015) also identified the same *Amh* variant in a Japanese strain of Nile tilapia originating from Egypt, which they named *AmhΔy* to distinguish it from another duplicated copy of the *Amh* gene, which they called *Amhy* due to its Y-specific expression and experimental evidence such as knockout the gene in XY and gene transfer into XX individuals (**Figure 4.1**). *Amhy* is located immediately downstream of *AmhΔy* and the coding sequence is identical to the X-linked *Amh* except for a 5608 bp promoter deletion and a missense SNP identified in exon II (thought to have a critical role in male sex determination). Li *et al*. (2015) proposed the *Amh* gene as a candidate sex-determining gene in Nile tilapia.

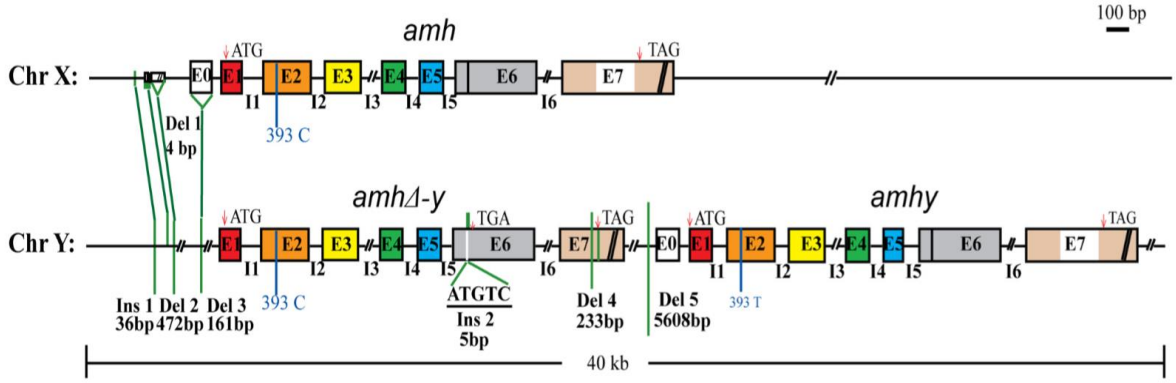


Figure 4.1 Gene structure of *Amh* (with seven exons) and Y-linked duplicated copies of *Amh: AmhΔy* and *Amhy*. *AmhΔy* has one insertion and three deletions in the upstream region, an insertion in exon VI and a deletion in exon VII. *Amhy* is tandemly located downstream of *AmhΔy*, lacks the promoter region and has a single base difference in exon II (adapted from Li *et al*. 2015).

Bulk segregant analysis (BSA) is a rapid mapping technique which was first proposed by Arnheim *et al.* (1985). The basic idea of BSA is that samples are pooled based on the phenotypic differences for a particular trait of interest and the genetic analysis then looks for differences between the pools. This technique has been previously used in mutation detection and disease studies in humans (Carmi *et al*. 1995; Amos *et al*. 2000); and genetic linkage study in plants (Michelmoore *et al*. 1991; Shen *et al*. 2003; Shashidhar *et al*. 2005; Yang *et al*. 2013). Subsequently the BSA approach has been combined with different molecular marker technologies to identify the QTL associated with disease resistance and sex-related markers in different fish species (Ezaz *et al.* 2004; Lee *et al.* 2004; Keyvanshokooh *et al.* 2007; Wang *et al.* 2009a, b; Wang *et al.* 2013). However, molecular marker development and genotyping needed to be done separately for most of the BSA-based marker analyses, which is costly and sometimes very time-consuming.

Over the last decade, NGS technologies have become widely used in discovering thousands of molecular markers in genomic research, and in sequencing whole genomes for model and non-model organisms. Because of the extremely large genome sizes and sequencing costs,

whole genome sequencing of individuals was initially limited to a small number of species, although as the technology has improved and costs have come down, this is now changing rapidly. An alternative approach to individual sequencing would be whole genome sequencing of pools of individuals (Futschik and Schlötterer 2010). Due to budget constraints, the cost for whole genome sequencing can also be reduced by sequencing targeted regions of the genome of pools of individuals (Van Tassell *et al.* 2008).

Restriction-site associated DNA (RAD) sequencing, a reduced representation sequencing strategy (a genotyping by sequencing technique), sequences the regions flanking the cut sites of a restriction enzyme in the target genome and is a powerful, relatively inexpensive tool to study genetic variation (Baird *et al.* 2008). RADseq has been used on pools of individuals to explore the potential use of BSA for SNP discovery and genome-wide population genetics, which compared to individual sequencing reduces the cost, time and labour many-fold (Baird *et al.* 2008; Schlötterer *et al.* 2014).

However, the genetic variability can be biased for analyses based on restriction digestion like RADseq. Allelic dropout will occur if the restriction site is polymorphic - such null alleles will make it impossible to genotype the associated SNPs (Gautier *et al.* 2013a; Schlötterer *et al.* 2014). For example about 9.4 % of heterozygous loci were found to be homozygote (due to allelic dropout) in the sequence data from a reduced representation library (RRL) in a human population genomics study when compared with the same individuals using a chip-based SNP genotyping assay (Luca *et al*. 2011). Another type of error can come from unequal representation of the individual genetic material which can increase the amplification errors in PCR-based NGS studies, thereby leading to variation in sequencing read depth (Akbari *et al.* 2005). These problems become more pronounced when RADseq is done on pooled samples because the individual genomes may not be

evenly sequenced in the pool; some genomes may be sequenced more frequently, which will lead to biased allele frequency estimates.

Although RADseq based on pooled genomes has challenges, it is possible to utilize the advantages of both RADseq and BSA by modifying the sequencing strategy, such as by increasing the number of individuals and adding equal amounts of DNA to make the pools. The sequencing error rates can also be minimized by using the powerful statistical software such as STACKS pipeline which was developed for RAD-based analysis (Gautier *et al.* 2013a, b; Schlötterer *et al.* 2014). For further reduction of cost and labour for DNA extraction and quantification, pre-extraction pooling of tissue samples followed by DNA extraction is another approach to RADseq for family-based association and population genetics studies which will allow sequencing of many samples at the same time.

Given the evidence for intraspecific and interspecific variation in sex determination in tilapia, and the synthetic base population from which GIFT was developed, this study began by applying a BSA version of ddRADseq to allow screening of many GIFT families for sex-determining loci, followed by more detailed screening of individual DNA polymorphisms.

4.3 Materials and methods

4.3.1 Sample collection and tissue preparation

GIFT families (from generation 19, personal communication John Benzie) were produced in WorldFish Center (Penang, Malaysia) as part of routine production of a new generation in the breeding programme, and reared under normal rearing conditions in family hapas until they reached the size at which fish for the breeding programme were removed and tagged.

The remaining fish were killed. Where sufficient fish remained to ascertain a meaningful sex-ratio, phenotypic sex was determined microscopically for each fish followed by preserving a fin clip in ethanol. Phenotypic data and fin tissue from 19 GIFT full sib families (progeny) and 50 broodstock (including the parents of the 19 families) were sent to the University of Stirling from the WorldFish Center and processed for further analysis. The phenotypic sex-ratio for each GIFT family used in this study is given in the **Appendix Table C4.1**.

In addition to the experimental animals (GIFT), two families from the Stirling Nile tilapia population were included as positive controls for the BSA-ddRAD analysis, after first verifying that the phenotypic sex-ratio was balanced and strongly associated only with SNP markers in LG1 (Palaiokostas *et al.* 2013a). The two Stirling Nile tilapia families were produced in the Tropical Aquarium Facilities, Institute of Aquaculture. Eggs and milt were stripped from ovulated female and male fish respectively and the eggs fertilised *in vitro* followed by incubation. Following yolk-sac absorption, larvae were transferred to the recirculatory system and reared for 3-4 months. After Schedule 1 killing, phenotypic sex (approximately 3 months old) was determined microscopically and a fin clip was fixed in 100 % ethanol for later DNA extraction.

Pre-extraction pooling of progeny fin tissue was done according to their phenotypic sex. An equal number of each sex was used to make the two pools (male and female progeny) for each family (**Table 4.1**). The number of individuals per family varied and depended on the numbers received from the WorldFish Center. For fin tissue preparation, an approximately equal amount of fin tissue from each individual was collected using a sterile 3 mm sized biopsy punch (Stiefel Laboratories Ltd) and was divided into two equal halves. One half of the tissue was used as a source of genomic DNA for each progeny and the other half was

returned to the tube with the rest of the fin tissue for that individual for later individual DNA extraction.

4.3.2 Genomic DNA extraction

Genomic DNA from individual samples was extracted using the salt precipitation method (modified from Aljanabi and Martinez 1997). Briefly, fin tissue was digested in lysis solution at 55^{o}C overnight. Following digestion, sample was treated with RNaseA at 37^{o}C for 60 min. Protein was precipitated by 5 M NaCl followed by DNA precipitation into absolute isopropanol. The DNA pellet was dissolved in 20 µL of 5 mM Tris (pH 8.0).

The same method was used for pooled samples, with some modifications. Pooled tissue (volumes given are for a pool of 30 individuals) was digested overnight at 55^{o}C in lysis solution containing 2.5 mL SSTNE/1 % SDS and 50 µL proteinase K (10 mg/mL). The digested tissue was allowed to cool followed by transferring 600 µL of the solution into a newly labelled Eppendorf tube. Following RNaseA treatment (30 µL RNaseA; 2 mg/mL) at 37^{o}C for 60 min, proteins were precipitated with 441 µL of 5 M NaCl (0.7 × vol). The supernatant (400 µL) was pipetted out into a newly labelled Eppendorf tube without touching the protein pellet, followed by DNA precipitation into 100 % ethanol (2.20 × vol). Following an overnight wash in 70 % ethanol, the DNA pellet was dissolved in 30 µL of 5 mM Tris (pH 8.0). The volumes described above (for 30 pooled individuals) were scaled down where smaller numbers of individuals were used in a pool (**Table 4.1**). Genomic DNA quantification, purity and integrity were assessed using spectrophotometry (Nanodrop, Labtech International Ltd, UK) and gel electrophoresis followed by diluting the DNA to 100 ng/µL with 5 mM Tris.

Table 4.1 Number of individuals used to make DNA pools for each family for BSA-ddRAD library construction.

ID	Strain	Number of individuals to make male progeny pool	Number of individuals to make female progeny pool
Family 1	Stirling	24	24
Family 2	Stirling	29	29
Family 1	GIFT	25	25
Family 2	GIFT	30	30
Family 3	GIFT	30	30
Family 4	GIFT	30	30
Family 5	GIFT	30	30
Family 6	GIFT	30	30
Family 7	GIFT	15	15
Family 8	GIFT	18	18
Family 9	GIFT	28	28
Family 10	GIFT	15	15
Family 11	GIFT	21	21
Family 12	GIFT	22	22
Family 13	GIFT	30	30
Family 14	GIFT	15	15
Family 15	GIFT	22	22
Family 16	GIFT	17	17
Family 17	GIFT	23	23
Family 18	GIFT	23	23
Family 19	GIFT	30	30

4.3.3 Bulk segregant analysis – double-digest restriction-site associated DNA (BSA-ddRAD) sequencing

4.3.3.1 Library preparation

The concentration of double-stranded DNA (dsDNA) was measured accurately using fluorimetry. Qubit® dsDNA broad range (BR) assay kits were used to measure the concentration of dsDNA using a Qubit® 2.0 Fluorometer (Life Technologies Ltd, UK) and after this each sample was diluted to a standard concentration of 8 ng/µL with 5 mM Tris (pH 8.0).

The first BSA-ddRAD library was constructed from 5 GIFT and 2 Stirling families as an initial test and each family consisted of dam, sire, male progeny pool and female progeny pool. Each individual sample (parental) was replicated three times and the pooled progeny samples were replicated four times, to generate more reads and as even sequencing coverage as possible across the genome. Once the first BSA-ddRADseq analysis showed that there was sufficient sequencing coverage (distributed throughout the genome and across all linkage groups) for subsequent analysis, the second library was constructed and sequenced for the rest of the GIFT families (n = 14) without replication during the library construction. The BSA-ddRAD library was prepared following the protocol described in Peterson *et al.* (2012) with some modifications. Each sample (24 ng DNA) was digested with *Sbf*I (CCTGCA^GG, rare cutter) and *Sph*I (GCATG^C, frequent cutter) high fidelity restriction enzymes (New England Biolabs, UK) at 37^{o}C for 90 min using 20 U of restriction enzyme per µg of genomic DNA in 10× CutSmart reaction buffer (New England Biolabs, NEB). Each digested DNA sample (total volume 6 µl) was ligated with individual-specific P1 (25 nM, *Sbf*I compatible) and P2 adapters (100 nM, *Sph*I compatible) for 2.5 hr at 22^{o}C, each with a unique 5 or 7 bp barcode, by adding 3 µl adapter mix (*Sbf*I:*Sph*I-1:12), 0.3 µl of 10× CutSmart reaction buffer, 0.12 µl of 100 mM rATP, 0.024 µl of 2 M ceU/ml T4 DNA Ligase and nuclease free water to make the final reaction volume 12 µl (barcode information in **Appendix Table C4.3**). Ligation was stopped by adding 2.5 volumes PB buffer (Qiagen, UK) and samples were multiplexed into one pool followed by purification with a single column (MinElute PCR purification kit, Qiagen, UK). The DNA fragments were then size selected and excised in the range of c. 400-700 bp on an 1.1 % agarose gel, followed by gel purification (MinElute gel purification kit, Qiagen UK) and then PCR amplification. Initially 13 and 16 cycles of PCR were used to determine the optimal PCR cyclic conditions for amplification of sufficient library for sequencing.

Finally the template was subjected to 11 cycles of PCR (using Q5 Hot-start High Fidelity DNA polymerase; NEB and Illumina specific primers) to obtain sufficient quantity (c. 400 µl) of amplified library for sequencing. The amplified library was purified in two steps: first by a column purification (MinElute PCR purification kit), followed by a paramagnetic bead clean up (AMPure XP, Beckman Coulter, UK), with a final elution into 20 µl warmed EB buffer. The purified amplicon library was quantified again using fluorimetry to measure the actual DNA concentration of the library for sequencing.

4.3.3.2 Preparation for library clustering and sequencing

The purified amplicon library was normalized to 10 nM stock with EB buffer and 1 % Tween 20. The stock amplicon library (10 nM) was diluted to 4 nM (2 µl of 10 nM library and 3 µl ddH_2O) followed by denaturation into 5 µl of fresh 0.2 M NaOH at room temperature for 5 min. The denatured library was diluted again into 20 pM library with 990 µl pre-chilled HT1 (hybridization buffer). The PhiX was used as an internal control during sequencing and was diluted to the same loading concentration (20 pM) as the amplicon library. The final concentration of the loading library was 10.2 pM; of which 5 % was PhiX (290.7 µl of 20 pM amplicon library, 15.3 µl of 20 pM PhiX library and 294.0 µl of HT1 buffer). The final library was heat denatured at 98^{o}C for 2 min, followed by cooling in an ice bath for 5 min, before loading on an Illumina MiSeq cartridge. Following the clustering procedure, the flow cell was loaded onto the MiSeq machine according to the manufacturer's instruction. The library was run using v2 sequencing chemistry in a single lane for 300 cycles (161 bases paired-end reads).

4.3.3.3 Computational methods for generating RAD loci

Raw sequence data (Fastq file format) were processed through the FASTQC software (Version 0.11.2), an initial quality control step for generated high-throughput sequence data. Multiplexed samples were demultiplexed using the default parameters of the process_radtags component in STACKS (version 1.27), a software pipeline designed for RAD-based analysis (Catchen *et al.* 2013). Low quality reads (process_radtags – s parameter set to 20), reads missing restriction enzyme cut sites and reads with ambiguous or unpaired barcodes were filtered out during demultiplexing. Filtered reads were aligned to the published Nile tilapia genome (Broad Institute of MIT and Harvard genome assembly Orenil1.1; Brawand *et al.* 2014) using the default parameters of the Bowtie 2 aligner (Langmead and Salzberg 2012) followed by formation of sequence alignment/map (SAM) files needed to build up RAD loci in STACKS. Reads were then sorted into loci using the default parameters of ref_map.pl component in the STACKS. The main STACKS parameters were set to m = 6, n = 1 for individual analysis of Stirling and GIFT families (m- minimum depth of coverage to build a stack in PSTACKS; n- mismatches allowed between catalog loci in CSTACKS). A multinomial-based likelihood model implemented in PSTACKS was used to evaluate each nucleotide position for each possible genotype, thereby identifying true SNPs from sequencing errors (Hohenlohe *et al.* 2010).

4.3.3.4 Genome-wide association studies to identify sex-linked markers in GIFT

Following the reference-based assembly within the STACKS, the resulting RAD loci were passed through the following filtering criteria to extract true loci for the downstream analysis: 1) RAD loci with at least one SNP (polymorphic) were considered, while all monomorphic RAD loci were removed; 2) RAD loci with more than two SNPs were discarded from the data; 3) common RAD loci for dam, sire, male progeny pool and female

progeny pool were needed for further analysis, so those having no data for either the parent or progeny pool were removed; 4) only bi-allelic loci for parents were included and poly-allelic loci were excluded; 5) the presence of both parental alleles in either (or both) progeny pools was ensured for the downstream analysis.

These filtering criteria were applied to ensure that only loci with all information needed for the subsequent analysis (criteria 1 and 3) and those being highly probable to reflect true polymorphisms, e.g. Mendelian segregation (criteria 2, 4 and 5), were taken into account. A Perl script was developed to fit with above filtering criteria and to filter out all the loci that did not fall into these criteria. Following filtration, a Fisher's exact test was performed between the datasets from the two progeny pools, using the exact nucleotide/allelic counts for the SNPs within the filtered loci, representing a first step of screening SNPs that might be significantly associated with phenotypic sex. The corrected *p*-values (*q*-value) were calculated using the qvalue/R package, a package that implements a false discovery rate (FDR) method for genome-wide tests of significance. To identify the positional candidate SNPs linked to sex, *q*-values were visualized according to the physical position in the Nile tilapia genome, using Manhattan plots in the qqman/R package (Turner 2014).

4.3.4 Further analysis of the identified sex-determining region in GIFT

Six GIFT families including parents and progenies (the three showing the highest sex-association with LG23 and the three showing the weakest or no significant association with LG23 from the BSA-ddRADseq analysis), all other GIFT broodstock (n = 12) and 4 Stirling broodstock (**Table 4.2**) were selected for further individual analysis. Five SNPs (showing the highest association with phenotypic sex in BSA-ddRADseq analysis) and two microsatellite markers (*UNH898, ARO172*) linked to sex in LG23 in the study of Eshel *et*

al. (2012) were genotyped (see section 4.3.4.1 and 4.3.4.2). The recently identified variations in Y-linked *Amh* (**Figure 4.1**) such as the 233 bp deletion in *Amh* exon VII (*Amhy*, Eshel *et al.* 2014; *Amh*Δ*y*, Li *et al.* 2015), 5 bp insertion in *Amh* exon VI (*Amh*Δ*y*, Li *et al.* 2015), exon 0 deletion in *Amh*Δ*y* (*Amh*_E0, Li *et al.* 2015), 5608 bp promoter deletion in *Amhy* (*Amhy*_Promoter_del, Li *et al.* 2015), 3 bp insertion in exon 0 *Amhy* (*Amh*_E0_del) and a single SNP (C/T) in *Amhy* exon II (*Amhy*_E0_E2, Li *et al.* 2015) were also screened. A missense SNP (G > C) in exon VI of *Amh* (*Amh*_SNP_exon_VI, Wessels *et al*. 2014) was also tested for GIFT individuals.

4.3.4.1 SNP marker genotyping

The two Stirling families (n = 110) were analysed using tightly sex-linked SNP markers in LG1 (SNPs *Oni23063* and *Oni28137*; Palaiokostas *et al.* 2013a) and in LG20 (*Oni3161*; Palaiokostas *et al.* 2015) to confirm the genotype-phenotype association before constructing the BSA-ddRAD library (**Table 4.1**). The five most highly significantly sex-associated SNPs from BSA-ddRADseq analysis in GIFT were selected and analysed for progenies of six GIFT families, and fifty GIFT and four Stirling broodstock (**Table 4.2**; NCBI dbSNP accession *ss2017360134*, *ss2017360168*, *ss2017360173*, *ss2017360175* and *ss2017360178*). A missense SNP (*ss831884014*) in exon VI of *Amh* was also tested for fifty GIFT and four Stirling broodstock (**Table 4.2**, Wessels *et al.* 2014). Allele-specific primers were designed based on c. 200 bases flanking these SNPs (Primer sequences are in **Appendix Table C4.4**).

SNPs were genotyped using fluorescence-based Kompetitive Allele Specific end-point PCR (KASP) genotyping system (KBioscience UK Ltd) following the protocol described in Semagn *et al*. (2014) with some modifications. The KASP master mix and allele-specific primers (KASP assay) were provided by LGC genomics (UK) based on supplied ddRAD

locus sequences. The PCR reaction volume was set to 5 µL (c. 25 ng DNA) and the PCR was performed using the following cyclic conditions: the initial denaturation at 94^{o}C for 15 min followed by 10 touchdown cycles (94^{o}C for 20 sec and touchdown 65^{o}C for 1 min, reduced by 0.8^{o}C per cycle) followed by 34 cycles of amplification (94^{o}C for 20 sec; 57^{o}C for 1 min). Amplified products were then analysed on a Techne Quantica® machine (Barloworld Scientific Ltd UK) to detect the fluorescence at ambient temperature. Allelic discrimination analysis was performed manually to determine the single nucleotide differences using the inbuilt Quansoft software (version 1.1.21).

4.3.4.2 Microsatellite marker genotyping

One microsatellite marker (*UNH995*) in LG1 was tested for two Stirling families (n = 110) and two microsatellite markers (*UNH898, ARO172*) in LG23 were analysed for six GIFT families, and fifty GIFT and four Stirling broodstock (n = 284, **Table 4.2**). Markers were analysed using the fluorescent labelled tailed primer method (Boutin-Ganache *et al.* 2001) with some modifications (primer sequences are in **Appendix Table C4.5**). In brief, 5 µL (c. 25 ng DNA) PCR reaction volumes was prepared and the thermal cyclic conditions were the initial denaturation at 95^{o}C for 1 min and 35 cycles of denaturation at 95^{o}C for 15 sec, annealing at 62^{o}C for 15 sec and extension at 72^{o}C for 30 sec. The labelled PCR fragments were then detected on a CEQ^{TM} 8800 capillary sequencer (Beckman Coulter®, USA) according to the manufacturer's instructions and the alleles were annotated using the default parameters in the CEQ software (version 9.0).

4.3.4.3 Variations in Y-linked *Amh*

The insertion in *Amh* exon VI and deletion in *Amh* exon VII were analysed for progenies in six families, and fifty GIFT and four Stirling broodstocks (n = 284, **Table 4.2**) using

standard PCR protocol and the amplified products (3 µL) were checked on 1.5 % agarose electrophoresis. The PCR was carried out in 5 µL reaction volumes (c. 25 ng DNA) and the thermal conditions were initial denaturation at 95^{o}C for 1 min and 35 cycles of denaturation at 95^{o}C for 15 sec, annealing at 62^{o}C (whereas 67^{o}C for 5 bp insertion in *Amh* exon VI) for 15 sec and extension at 72^{o}C for 30 sec. The promoter deletion (5608 bp) in *Amhy* was also checked for five GIFT individuals (one female and four males). A 30 µL PCR reaction volume was prepared with TaKaRa LA Taq® Hot Start DNA polymerase (5 units per µL, Takara Bio Europe SAS, UK). PCR cyclic conditions were the initial denaturation at 96^{o}C for 2 min, 30 cycles of denaturation at 96^{o}C for 40 sec, annealing at 63^{o}C for 30 sec and extension at 72^{o}C for 8 min 30 sec, with the final extension at 72^{o}C for 10 min. The amplified product (2 µL) was checked on a 1.5 % agarose gel. The exon 0 deletion in *Amh∆y* (*Amh*_E0) and 3 bp insertion in *Amhy* exon 0 (*Amh*_E0_del, primers were designed upstream and downstream of the exon 0 region, primer sequences are in **Appendix Table C4.5**) were analysed for fifty GIFT and four Stirling broodstock (n = 54). PCR was carried out with 64^{o}C annealing temperature, and following PCR *Amh*_E0 amplified products were checked on 1.5 % agarose gel, and *Amh*_E0_del amplified products were run on a CEQ genotyping machine for fragment analysis. A primer pair was designed from the middle of exon 0 to beyond exon II in *Amhy* (1234 bp) to verify the SNP in *Amhy* exon II in GIFT individuals. Eleven GIFT sires and one dam were processed for Sanger sequencing (1234 bp) following the GATC Biotech manufacturer's instruction (Sanger ABI 3730×l, LIGHTRUN™ sequencing service). Another primer pair was designed flanking a SNP (*ss831884014*, Wessels *et al.* 2014) in exon VI of *Amh* and was processed for six individuals (2 dams and 2 sires from GIFT, and the dam and sire for Stirling family 1) for Sanger sequencing following the GATC Biotech manufacturer's instruction. Primer sequences for all the markers tested are in **Appendix Table C4.5**.

Table 4.2 GIFT and Stirling fish analysed for individual SNP and microsatellite markers, and sequences in LG23.

ID	Number of male progeny	Number of female progeny	Total number of fish analysed
Family 1	20	20	40
Family 2	20	20	40
Family 3	20	20	40
Family 4	20	20	40
Family 7	15	15	30
Family 19	20	20	40
Broodstock (GIFT)	-	-	50
Broodstock (Stirling)	-	-	4

4.3.5 Association analysis between DNA marker and phenotypic sex in GIFT

An association analysis between genotype and phenotypic sex for each LG23 SNP marker was conducted for each family and broodstock using the SNPassoc package in R (version 3.1.3). A generalized linear model was applied under the function *WGassociation* to test the magnitude of association between each SNP marker and phenotypic sex. Significant *p*-values were corrected for multiple tests using the Bonferroni correction method. Fisher's exact test was used to test for significance for association of microsatellite and *Amh* gene variations (located on chromosome Y) to the phenotypic sex.

Data access

The raw sequence data for this study were submitted to the EBI's European Nucleotide Archive (ENA) Sequence Read Archive (SRA), study accession number PRJEB13792. Five SNPs were submitted to dbSNP NCBI and the accession numbers are *ss2017360134*, *ss2017360168*, *ss2017360173*, *ss2017360175* and *ss2017360178*.

4.4 Results

4.4.1 Sex-linked marker analysis in Stirling family

The Stirling families were analysed using tightly sex-linked markers in LG1 - SNPs (*Oni23063* and *Oni28137*) and microsatellite (*UNH995*) - and also a SNP marker in LG20 (*Oni3161*), to confirm the genotype-phenotype association before constructing the BSA-ddRAD library. The phenotypic sex-ratios in the two Stirling families were not significantly different from the expected 1:1 ratio (*p*-values of 0.777 and 0.999 for families 1 and 2 respectively). SNP markers (*Oni23063* and *Oni28137*) in LG1 showed strong, significant association with sex for both of the families (*p*-value of 2.606e-33 for both SNP markers), as did LG1 microsatellite marker (*UNH995, p*-value = 8.227e-15 and 2.129e-14 for families 1 and 2 respectively). Only one male and female (out of 59) from the second family did not follow the expected genotype segregation pattern with respect to their phenotypic sex, for *UNH995* (**Table 4.3**; **Appendix Table C4.2**). In contrast, the genotype-phenotype association analysis for the SNP marker (*Oni3161*) in LG20 was not significant in either case, suggesting that there was no influence of this marker to the phenotypic sex in either family (**Table 4.3**). On this basis, these two Stirling Nile tilapia families were used as positive controls for BSA-ddRADseq.

Table 4.3 Allelic segregation for SNP markers (*Oni23063* and *Oni28137*) and microsatellite marker (*UNH995*) in LG1, and SNP marker in LG20 (*Oni3161*) for the offspring in Stirling families 1 and 2.

		SNPs				***UNH995***			
		LG1: G/G (*Oni23063*) or T/T (*Oni28137*) Female expected genotype		LG1: A/G (*Oni23063*) or G/T (*Oni28137*) Male expected genotype					Fisher's exact test (*p*-value)
		Male	Female	Male	Female	Genotype	Male	Female	
Family 1 (*Oni3161*)	LG20 C/T	0	10	13	0	188/188	14	0	
	LG20 C/C	0	14	13	0	188/190	12	0	8.227e-15
						184/188	0	11	
						184/190	0	13	
Family 2 (*Oni3161*)	LG20 T/T	0	5	6	0	188/188	18	0	
	LG20 C/T	0	19	13	0	184/188	10	1	2.129e-14
	LG20 C/C	0	5	10	0	184/192	1	14	
						188/192	0	14	

4.4.2 BSA-ddRAD sequencing

In total, 57,012,594 raw reads (28,506,297 paired-end reads) of 161 bases were generated from the two sequencing runs. At the first step of quality filtering, 83.57 % of the raw reads (47,645,454; 23,822,727 paired-end reads) were retained after removing the low quality reads, reads with no restriction enzyme cut site and ambiguous or unpaired barcodes. Following filtering, reads were aligned to the published Nile tilapia genome using the STACKS package for each family separately to build the BSA-ddRAD loci.

4.4.3 Generating RAD loci in Stirling and GIFT families

After aligning the filtered reads from the first (triplicated and quadruplicated samples) sequencing run to the Nile tilapia genome, > 13 K to 15 K and > 15 K to 16 K RAD loci were generated from the two Stirling and five GIFT families respectively from the first BSA-ddRAD sequencing run. In the second sequencing run, > 9 K to 13 K RAD loci were generated from the other 14 GIFT families (**Table 4.4**). The number of RAD loci for the families in each sequencing run was mostly similar and the average read coverage per locus per DNA sample was c. 48× (minimum c. 25× and maximum c. 97×).

Bi-allelic polymorphic loci with no more than two SNPs and showing Mendelian inheritance were selected for the downstream analysis to identify the sex-determining region(s) in all the families studied. The loci that did not follow those criteria were filtered out from the generated RAD loci (c. about 80 %). After the final filtration step, > 1 K to 2 K and > 1 K to 3 K informative SNPs were kept for two Stirling and 19 GIFT families respectively for subsequent association analysis (**Table 4.4**).

Table 4.4 Details of the number of reads before and after initial filters following sequencing and the number of RAD loci generated using Nile tilapia genome and final filtered polymorphic loci used for downstream analysis.

Family	Source	Run	Raw reads	Initial filtered total reads	Initial filtered paired-end reads	RAD loci	Final filtered polymorphic loci
Family1	Stirling	1st	31,635,640 (15,817,820 paired-end reads)	3,663,294	1,831,647	15,757	2,519
Family2				2,879,600	1,439,800	13,724	1,432
Family1	GIFT			3,259,182	1,629,591	16,064	2,880
Family2				4,339,922	2,169,961	16,667	3,402
Family3				4,357,524	2,178,762	16,711	3,342
Family4				3,463,356	1,731,678	15,783	3,133
Family5				3,675,630	1,837,815	16,211	2,989
Sub-Total				**25,638,508**	**12,819,254**	**110,917**	**19,697**
Family6	GIFT	2nd	25,376,954 (12,688,477 paired-end reads)	1,287,394	643,697	9,948	1,744
Family7				1,511,434	755,717	10,865	2,034
Family8				1,186,308	593,154	10,274	1,828
Family9				1,328,738	664,369	10,449	1,816
Family10				1,752,852	876,426	11,392	2,177
Family11				1,701,066	850,533	11,673	2,246
Family12				1,452,378	726,189	10,461	2,023
Family13				1,419,650	709,825	10,881	1,875
Family14				1,572,212	786,106	11,344	2,059
Family15				1,669,660	834,830	11,066	1,959
Family16				1,525,696	762,848	11,440	1,979
Family17				2,727,446	1,363,723	13,364	2,300
Family18				1,540,058	770,029	11,168	2,127
Family19				1,332,054	666,027	10,573	1,907
Sub-Total				**22,006,946**	**11,003,473**	**154,898**	**28,074**

4.4.4 Mapping of sex-linked region from BSA-ddRADseq analysis

To identify the genomic region linked to phenotypic sex, an association analysis was performed for each family separately with the exact nucleotide/allelic counts for the filtered SNPs in progeny pooled data. The sex-linkage probability (*q*-value) for each SNP was plotted against its physical position corresponding to the published Nile tilapia reference

genome. In the case of the Stirling families, SNP markers having high probabilities of being linked to phenotypic sex clustered in chromosome one, as expected (**Figure 4.2A**). No other chromosomes (LGs) with high linkage probability to phenotypic sex were detected. BSA-ddRADseq based tilapia sex determination analysis confirmed the major sex-determining locus (LG1) in the two Stirling families, which gave similar results in sex-linked marker analysis for those families before constructing BSA-ddRAD library and also matched with the previous mapping results obtained from individual sequencing (using standard RADseq) in the same tilapia stock (Palaiokostas *et al.* 2013a).

On the other hand, SNPs with high linkage probability to the phenotypic sex clustered in a different chromosome (LG23) in the GIFT families (**Figure 4.2B**). No other significant associations appeared across the rest of the genome. A strong significant association was found in 12 GIFT families, while four families showed weaker but significant association, with some "noise" in the lower part of the graphs (Families 1, 2, 6 and 10; **Appendix Figure C4.1**) and three families did not show any significant association (Families 5, 14 and 19; **Appendix Figure C4.1**). Pre-extraction pooling of fin tissue followed by sequencing uncovered the location of the putative sex-determining region in the GIFT population (developed from a mixture of stocks), which was previously unknown.

The physical position in the genome of the five most significantly sex-associated SNPs (from BSA-ddRAD analysis) across all the families was then determined to narrow down the position of the sex-linked region of LG23. These SNPs were found at the physical positions of 8-13 Mb in LG23 (**Figure 4.4A**), and the identified region included the *Amh* gene and the two sex-linked microsatellite (*UNH898*, *ARO172*) markers (Eshel *et al*. 2012, 2014; Li *et al*. 2015).

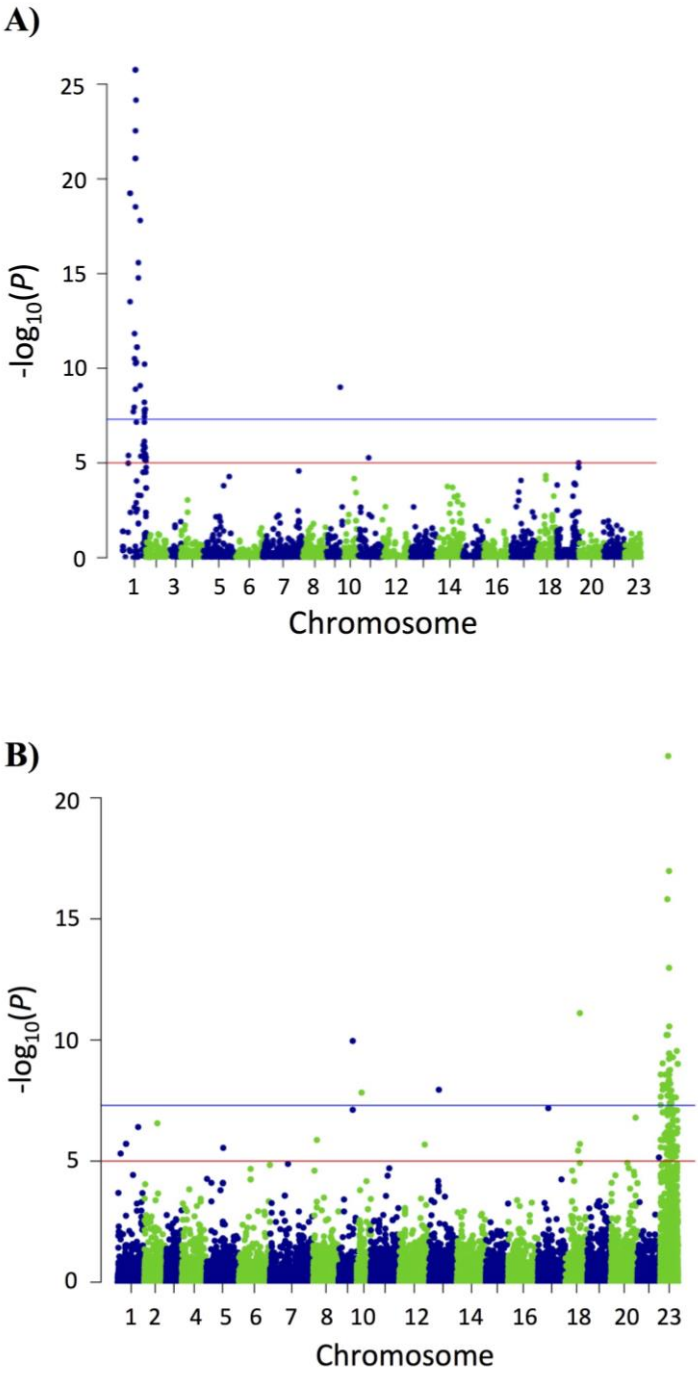


Figure 4.2 Genome-wide association plots, from combined families of Stirling and GIFT. Each dot represents a SNP and the Y-axis represents the magnitude of association ($-\log_{10}P$ value of F-test) of the SNP with phenotypic sex, while the X-axis represents the position in the linkage groups of the assembled Nile tilapia genome. The alternating blue and green colours are used to distinguish between chromosomes. The red solid line represents a *q*-value of 0.05 and the blue solid line represents a *q*-value of 0.01. A) SNPs significantly associated with the phenotypic sex were identified in LG1 for Stirling families B) SNPs in LG23 showed highest significant association with the phenotypic sex in GIFT families.

4.4.5 Association analysis between genotype and phenotype at the family and population level, from individual marker assays

In total fourteen LG23 DNA markers (5 SNPs from BSA-ddRAD analysis; 2 microsatellites; 2 insertions, 3 deletions and 2 SNPs in the variation of *Amh* gene) were analysed for 6 GIFT families (230 progeny in total), 50 GIFT and 4 Stirling broodstock (**Table 4.2**). The markers tested showed strongest association with the phenotypic sex in GIFT families and broodstock, but some deviations between the genotypic and phenotypic data were found to be common for all the markers tested for some progenies in four GIFT families and 2 broodstock analysed (**Table 4.7**). The results also suggest that the fin tissues of parents from GIFT family one were transposed (**Table 4.7**).

The parents for six GIFT families studied were not informative for some of the SNP markers, therefore the SNP markers were analysed individually only for the families where the parents were informative for the particular SNP (**Table 4.5**). The *p*-value threshold (0.05) was corrected to 0.01 taking into account the multiple tests performed (Bonferroni correction). All the five SNP markers were found to be significantly associated with phenotypic sex in the GIFT families where the sire was heterozygous (informative) for each SNP (**Table 4.5**). In case of SNP markers *ss2017360134*, *ss2017360173* and *ss2017360178*, the allele found in the homozygote was always associated with female and allele found only in the heterozygote was always associated with male where the dam was homozygous and the sire was heterozygous for those markers, except for 5 male and 3 female progeny (in total 310 genotype data). In the case of family 4, where both parents were heterozygous for the *ss2017360173* marker, there was no deviation of the genotype segregation among the progeny from the expected ratio 1:2:1 (*p*-value 0.752; **Appendix Table C4.6**). The T from the male parent was associated with male progeny, while the C from the male parent

was associated with female progeny; the female progeny were a mixture of CC and CT, while the male progeny were a mixture of CT and TT. In case of *ss2017360168* and *ss2017360175* markers, the results suggested that recombination might have occurred in all the sire informative families. Therefore the T allele from the male parent was always associated with female progeny (CT) and the C allele from male parent was associated with male progeny (CC) in the case of the *ss2017360168* marker; whereas for *ss2017360175* marker the T allele from male parent was associated with female progeny (AT) and the A allele from the male parent was linked to the male progeny (AA), with three exceptions for each of these two markers (in total 150 genotype data, **Appendix Table C4.6**). The physical position of those 5 SNPs were in a region of c. 3 Mb in the assembled Nile tilapia genome (**Figure 4.4A**).

In the case of broodstock, two of the five SNP markers (*ss2017360173* and *ss2017360175*) showed significant association with the phenotypic sex (*p*-values were 3.96e-05 and 3.91e-03 respectively, **Table 4.5**). Female broodstock were homozygous and males were heterozygous for the marker *ss2017360173*, apart from six males and four females (out of 50, **Appendix Table C4.6**). Progeny sex data were available for all these broodstocks. Two of the six males were thought to be sex-reversed; one of them (crossed with an XX female) gave nearly all female progeny (98.67 %) suggesting that the male was an XX neo-male, while the other male (crossed with an XX female) gave a progeny sex-ratio highly skewed to males (89.83 %) suggesting that the sire was YY. The marker association study strongly suggested that the fin tissues for the one male and one female (the parents of family 1) had been transposed during collection/labelling. For another male, progeny information was available but there was no dam sample to allow analysis and any conclusions. A few more mismatches were observed in the case of the *ss2017360175* marker with the phenotypic sex. These two SNPs are c. 227 K bases apart in the published tilapia genome. The results from

the broodstock analysis show that the *ss2017360173* marker was the most closely sex-linked marker in the GIFT populations among the SNPs derived from BSA-ddRADseq analysis.

Analysis of the SNP (*ss831884014*, Wessels *et al.* 2014) in exon VI of *Amh* analysis for GIFT and Stirling broodstock did not reveal any polymorphism, so it was not possible to draw any conclusions about this.

Table 4.5 Association analysis between phenotypic sex and each SNP marker for 6 GIFT families and 50 GIFT broodstock with their respective *p*-values.

		SNP markers									
		ss2017 360134	***P*-value**	***ss2017 360168***	***P*-value**	***ss2017 360173***	***P*-value**	***ss2017 360175***	***P*-value**	***ss2017 360178***	***P*-value**
Family 1	Dam	C/C	N/A	C/C	0.751	T/T	N/A	A/A	N/A	G/G	N/A
	Sire	C/C		C/T		C/C		A/A		A/A	
Family 2	Dam	C/C	6.76e-05	C/C	2.259e-05	C/C	6.76e-05	T/T	N/A	G/G	N/A
	Sire	C/T		C/T		C/T		A/A		G/G	
Family 3	Dam	C/C	8.756e-07	C/C	1.083e-06	C/C	1.145e-08	A/A	N/A	A/G	0.999
	Sire	C/T		C/T		C/T		A/A		G/G	
Family 4	Dam	C/C	N/A	C/C	N/A	C/T	1.00e-05	A/T	0.757	A/G	0.757
	Sire	C/C		C/C		C/T		A/A		A/A	
Family 7	Dam	C/C	N/A	C/C	N/A	C/C	5.656e-08	A/A	6.70e-07	A/G	0.694
	Sire	C/C		C/C		C/T		A/T		G/G	
Family 19	Dam	C/C	1.61e-03	C/C	4.76e-03	C/C	3.15e-05	A/A	N/A	G/G	1.31e-04
	Sire	C/T		C/T		C/T		A/A		A/G	
	Brood stock		0.111		0.439		3.96e-05		3.91e-03		0.356

The allelic distributions of the two LG23 microsatellite markers (*UNH898* and *ARO172*) were found to be significantly associated with the phenotypic sex when tested across all the families (*p*-value < 2.2e-16 for each marker) and each family separately for each marker (**Table 4.6**). The observed genotype frequency for these two microsatellite markers in all the families did not deviate from the expected ratio. For the population data (50 broodstock), these two microsatellite markers were also highly associated with the phenotypic sex (*p*-value 7.62e-06 and 6.54e-07 respectively). The 267 allele for *UNH898* and 274 allele for *ARO172* marker were nearly always associated the male phenotype, irrespective of family and broodstock, with few exceptions (**Table 4.7; Appendix Table C4.6**). In the case of two GIFT families (no. 4 and 7), the 267 and 274 alleles for *UNH898* and *ARO172* respectively were 100 % associated with males, with no deviation. The individual concluded to be an XX male did not have either Y-linked allele and the individual concluded to be a YY male was homozygous for the Y-linked allele for both markers (**Appendix Table C4.6**; ID 20S and 22S). The microsatellite marker data also suggested that the tissue samples for the parents of family 1 were wrongly labelled.

PCR-based markers (*Amh* exon VI, exon VII and *Amh*_E0) were also found to be significantly associated with phenotypic sex in all the GIFT families and broodstock tested (**Table 4.6**; **Figure 4.3**). A 233 bp deletion for *Amh* exon VII and 547 bp band for *Amh* exon VI were found to be nearly always associated with the male phenotype, with the same exceptions as found for the other markers mentioned above (**Figure 4.3A, B**; **Appendix Figure C4.2**; **Table 4.6, 4.7**; **Appendix Table C4.6**). Both *Amhy* (439 bp) and *AmhΔy* (233 bp deletion) linked bands were evident in the putative YY male (**Figure 4.3A**, ID 22S). A 233 bp deletion and 547 bp band for *Amh* exon VII and exon VI respectively were always associated with the male phenotype in two GIFT families (no. 4 and 7). The 161 bp deletion (*Amh*_E0) in *AmhΔy* (Li *et al*. 2015) was observed in nearly all male individuals with the

same exceptions found for the other markers (**Figure 4.3C**). The YY individual in **Figure 4.3C** (ID 22S) showed only *Amhy*-specific band for *Amh*_E0 (Li *et al*. 2015) and following sequencing of that region, polymorphisms were identified in one of the primer binding sites. The data from all of these markers supported the conclusion that the parents of family 1 were wrongly labelled. A 5608 bp promoter deletion was also observed in the XY and YY GIFT males studied (**Figure 4.4C)**.

Table 4.6 Association analysis of microsatellites and markers in the variations of the *Amh* gene on the Y chromosome with the phenotypic sex of progeny for each family separately and broodstock.

	Microsatellite markers		**Markers from the variation of *Amh* gene on Y chromosome**	
ID	*UNH898*	*ARO172*	*Amh* exon VI	*Amh* exon VII
Family 1	5.83e-05	5.83e-05	1.00e-05	1.00e-05
Family 2	1.96e-04	1.96e-04	1.37e-04	1.37e-04
Family 3	1.60e-06	3.36e-06	5.30e-07	5.30e-07
Family 4	4.78e-08	2.15e-08	1.45e-11	1.45e-11
Family 7	1.52e-08	1.79e-08	1.29e-08	1.29e-08
Family 19	9.27e-04	1.28e-03	3.60e-04	3.60e-04
Broodstock	7.62e-06	6.54e-07	5.21e-10	5.21e-10

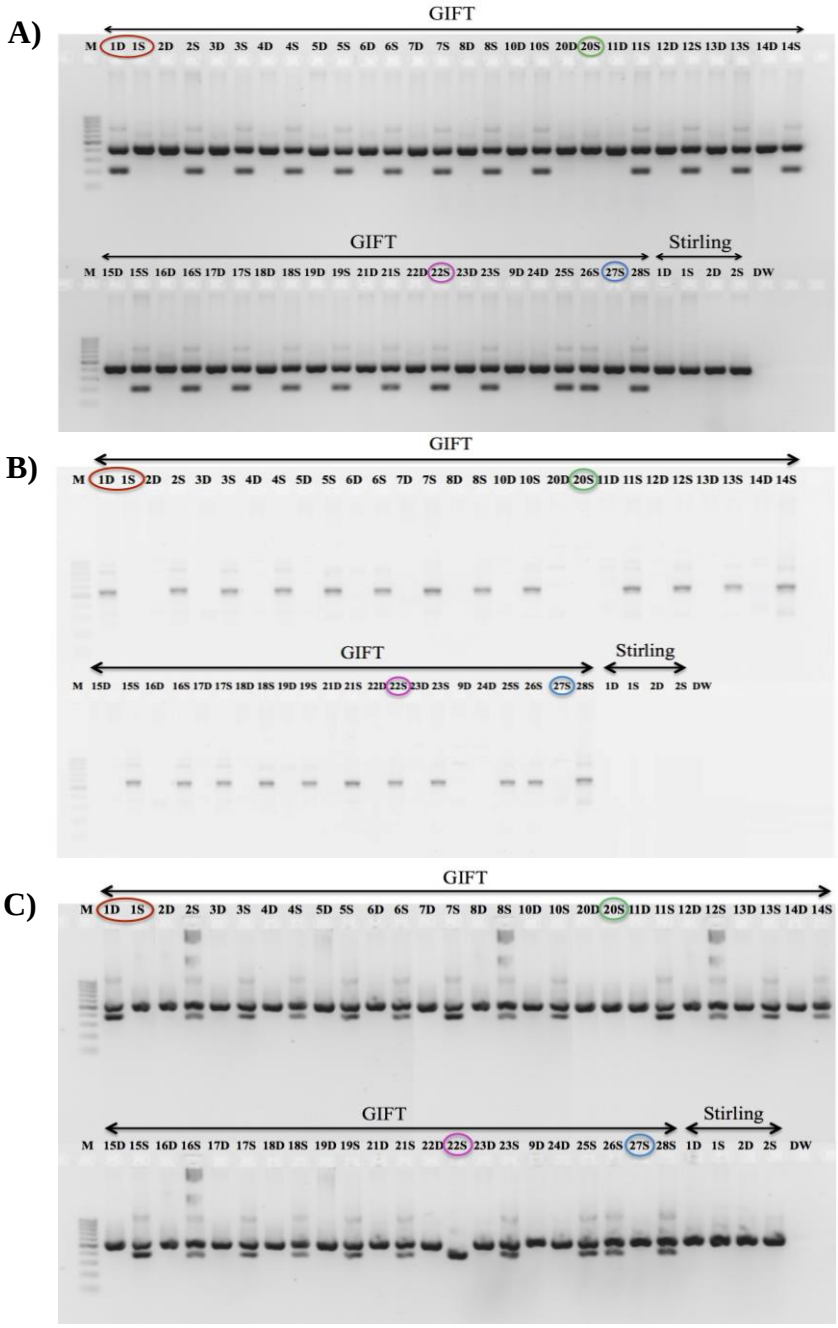


Figure 4.3 Amplified PCR products on 1.5 % agarose gel using markers from the different regions in the variations of the *Amh* gene for 50 GIFT and 4 Stirling broodstock. A) Deletion in *Amh* exon VII (*Amhy*, Eshel *et al*. 2014). A 439 bp band was evident in all individuals and a 233 bp deletion was present in nearly all males. B) Insertion in *Amh* exon VI (*AmhΔy*, Li *et al*. 2015); a 547 bp band was present in nearly all sires and dams showed no band. C) Exon 0 deletion (161 bp) in *AmhΔy* (Li *et al*. 2015); a 547 bp band was present in all individuals except 22S (purple circle, putative YY) and nearly all males showed a band with 386 bp (had 161 bp deletion). The exceptions are - red circle indicates dam and sire concluded to be wrongly labelled (transposed) individuals, green circle indicates sire concluded to be an XX neo-male, progeny information was available for blue circled sire but the dam information was not available. M - molecular marker (100 bp), D - dam, S – sire, DW - distilled water. Stirling broodstocks were non-informative.

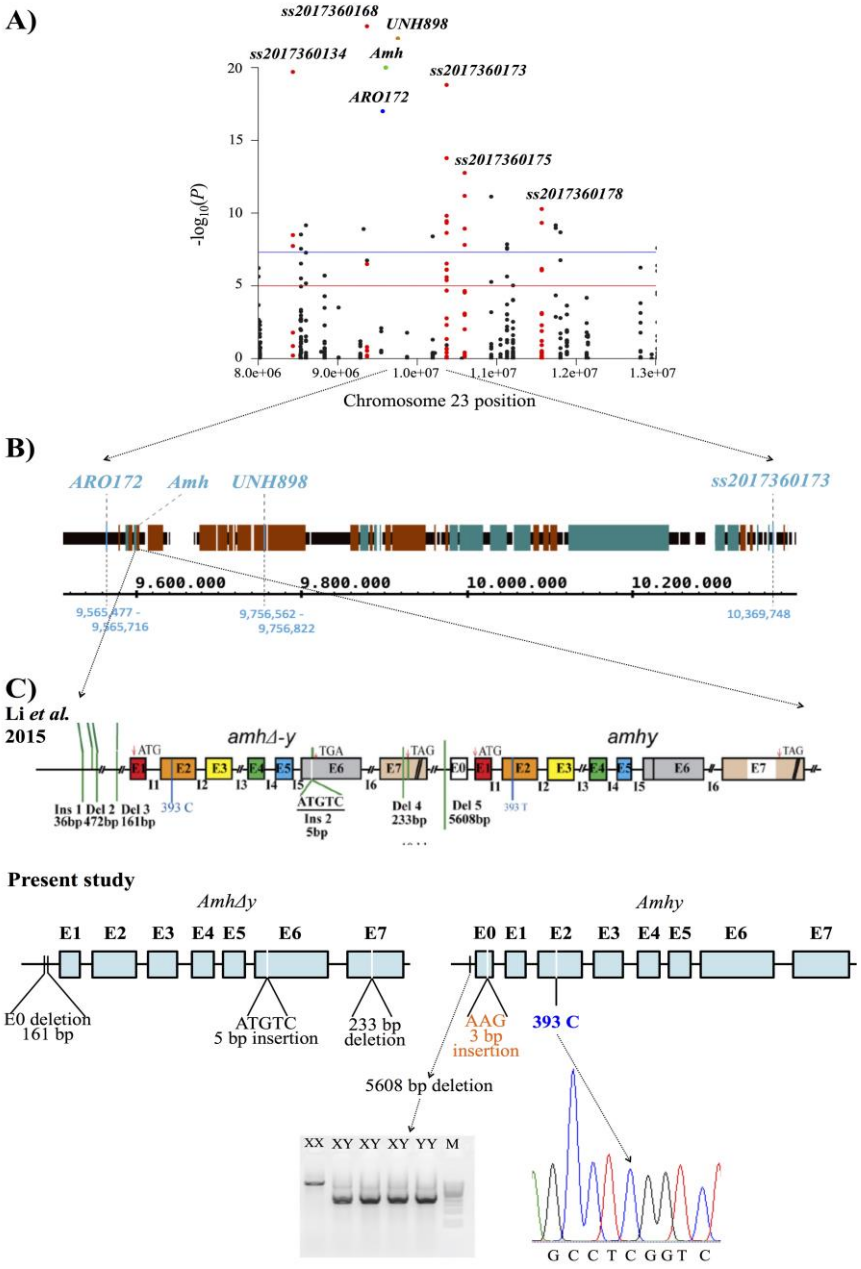


Figure 4.4 Detail diagram of the putative XX/XY sex-determining region in LG23 in GIFT. A) Position of the 5 SNPs, 2 microsatellite markers and *Amh* gene along the assembled Nile tilapia chromosome 23. Each dot represents the magnitude of association between the particular SNP and phenotypic sex for each family. B) Gene information about the region of higher association (position 9,560,000 to 10,370,000). It includes 32 genes of which 14 are annotated, with 26 gaps (19 to 29,961 nt; white regions). Green: genes on the plus strand, red: genes on the minus strand, black: normal nts with no identified gene. C) Similarities and dissimilarities in the analysed Y-linked *AmhΔy* and *Amhy* between GIFT and Li *et al*. (2015); *AmhΔy* had exon 0 deletion (161 bp), 5 bp insertion in exon VI (ATGTC) and 233 bp deletion in exon VII; *Amhy* had 5608 bp promoter deletion and 3 bp insertion (AAG) in exon 0 region and did not have polymorphism in exon II (C/C) in the GIFT individuals studied.

A three base pair insertion was also identified in exon 0 in *Amhy*, compared to the sequence published by Li *et al*. (2015). Males showed this three base pair insertion, while females did not, with three exceptions - two males and one female. These males had the 253/256 genotype whereas the female had the 253/253 genotype (**Appendix Table C4.6**). Eleven GIFT sires and one dam were sequenced to test for the polymorphism (C/T) in the exon II described by Li *et al*. (2015) and thought by these authors to have a critical role for male sex-determination in Nile tilapia. No such polymorphism was detected in the GIFT individuals studied - all had the base C in this position. Sequence analysis for another sex-linked SNP (*ss831884014*, Wessels *et al.* 2014) in exon VI of *Amh* did not show polymorphism at this site in the six samples analysed (2 dams and 2 sires from GIFT, and the dam and sire for Stirling family 1).

Based on the individual analysis of 6 GIFT families, some individuals showed departures of markers from the expected phenotypic sex; these were common for all the markers tested (**Appendix Table C4.6**). In the cases of families 1 and 2, five and seven phenotypic males respectively had the female expected genotype, whereas only one female from each family had the male expected genotype. The association between SNPs in LG23 and phenotypic sex in BSA-ddRAD analysis was very weak for those families (**Appendix Figure C4.1**). In case of family 3, very small but equal numbers i.e. two males and two females came up with the female and male genotype respectively, and a very strong association was found between SNPs and sex in BSA-ddRADseq analysis. On the other hand, slightly higher but equal numbers of males (n = 4) and females (n = 4) had the female and male expected genotype respectively in case of family 19, which did not show significant association in any of the genomic regions in BSA-ddRADseq analysis.

Stirling broodstock (4 samples) were non-informative for all the SNP markers in LG23 analysed and no male-associated microsatellite alleles or Y-linked bands (in LG23) were evident in those broodstock (**Figure 4.3**; **Appendix Table C4.6**).

Table 4.7 Agreement and disagreement of phenotypic sex and genotype segregation for each family and broodstock studied for five SNPs, two microsatellites markers and markers in the variation of *Amh* gene (deletion in *Amh* exon VII and insertion in *Amh* exon VI) in Y chromosome.

ID	Phenotype	Female expected genotype	Male expected genotype	Total number
Family 1	Female	19	1	40
	Male	5	15	
Family 2	Female	19	1	40
	Male	7	13	
Family 3	Female	18	2	40
	Male	2	18	
Family 4	Female	20	0	40
	Male	0	20	
Family 7	Female	15	0	30
	Male	0	15	
Family 19	Female	16	4	40
	Male	4	16	
Broodstock	Dam	23	1	50
	Sire	3	23	

Insertion and deletions in *AmhΔy* were found to be strongly associated with the phenotypic sex in the GIFT samples analysed and were the same as was found in the Japanese population of Nile tilapia (Li *et al*. 2015). In contrast, the variations in *Amhy* were different

in GIFT from the Japanese population (Li *et al*. 2015). A three base pair insertion in exon 0 region in *Amhy* was found to be linked to the male sex determination in GIFT (no information in Li *et al*. 2015).

A single SNP marker (*ss2017360173*) from the BSA-ddRAD analysis, two microsatellite markers (*UNH898, ARO172*) and the markers in the variations of the *Amh* gene (*Amhy* and *Amh∆y*) showed the strongest association with phenotypic sex in GIFT and all of these markers are located in a range of c. 804 K bases in the Nile tilapia genome (**Figure 4.4A**). The 14 annotated genes are located within this region include the *Amh* gene (**Figure 4.4B**) which suggests that the Y-linked *Amh* gene(s) could be responsible for the male sex determination in GIFT, as suggested by Eshel *et al*. (2014) and Li *et al*. (2015).

4.5 Discussion

In this study, we developed a powerful extension of BSA and ddRADseq by applying pre-extraction pooling of tissue samples to ddRADseq for the analysis and identification of sex-determining region(s) in GIFT, followed by the verification of the identified region with different molecular marker analyses. This quick mapping technique allowed rapid analysis at the population level, location and identification of the major sex-determining region in GIFT. The markers from this region could be applied in marker-assisted selection for the first time for controlling sex-ratio in GIFT in the culture systems.

4.5.1 BSA-ddRADseq is a useful tool for family-based association analysis

The DNA pooling strategy was originally used with standard molecular techniques to identify markers linked to genes or genomic regions of interest (Giovannoni *et al*. 1991; Michelmore *et al*. 1991; Wang and Paterson 1994). With the rapid advancement of NGS

technologies, BSA has been incorporated with different NGS platforms given the importance of sequencing lots of individuals with a minimum cost, although there are possibilities for errors, mentioned previously for the BSA technique, and these can be more pronounced when using BSA combined NGS (Baird *et al*. 2008; Wang *et al*. 2013; Ramirez-Gonzalez *et al*. 2015).

The variation can come from low number of individuals in each bulk and heterogeneity in the DNA amounts added into the bulk. A significant difference in the contribution of DNA per sample in the pool may result bias or suboptimal results for allele frequency estimates of individual SNPs. This unequal representation of individuals does not have severe effects on the accuracy of allele frequency estimates unless the sample size is very small. This type of error can be minimized by adding approximately equal amount of genetic material to the pool and by increasing the number of individuals in the pool (Ferretti *et al*. 2013; Futschik and Schlötterer 2010).

In prior publications the genetic material was extracted individually followed by pooling of approximately equal amounts of genetic material. Pre-extraction pooling of tissue samples before DNA extraction for NGS has been applied on a limited scale in plants (Ramirez-Gonzalez *et al*. 2015) and invertebrates (Bastide *et al*. 2013); no reports were found on the pooling of vertebrate animal tissue samples and the challenges of using this for NGS techniques. As tissue samples were pooled together in the present study, it was thought that there might be more variation in the representation of the genetic material from each individual. However, this strategy reduced the extraction and sequencing cost, time and labour and allowed analysing of hundreds of individuals from a single population at the same time.

In genotype-phenotype association studies, pools have been constructed with extreme phenotypic differences to screen loci linked to the quantitative trait. In family based-association studies, allele frequencies or allelic associations are compared between the parent and the segregated progeny (Kirov *et al*. 2000). In the present study, considering the variation likely to be present within the pooled progeny samples, only bi-allelic polymorphic loci showed Mendelian inheritance were used for the association analysis. In the two Stirling families a clear, strong association signal was identified as expected between phenotypic sex and LG1 markers from the BSA-ddRAD analysis, as the same association was found using known LG1 markers before constructing the BSA-ddRAD library and in the previous study with the same stock (Palaiokostas *et al*. 2013a). On the other hand, a strong association was found in a different chromosome (LG23) in the majority of the GIFT families.

Different study reported different number of individuals required to construct the pool for population genetics or genotype-phenotype association study. For example Schlötterer *et al*. (2014) reported that BSA with NGS performs well when 50 individuals are pooled and larger pools than this (> 100) can result in even higher accuracy in allele frequency estimates. Such large numbers are sometimes not feasible for some species, e.g. endangered ones. Another report has mentioned that ≥ 50 individuals for haploid organisms or > 20 individuals for diploid organisms in a pool for NGS would have the power to estimate allele frequency accurately (Ferretti *et al*. 2013). It has also been suggested that 10 to 20 individuals in each pool are sufficient to screen markers affecting a specific trait of interest, for example candidate gene mapping, QTL mapping and SNP marker discovery (Livaja *et al*. 2013; Randhawa *et al*. 2014; Wang *et al*. 2014). Given the evidence on different numbers of individuals in such pools, our study was designed with a range of individuals (15 to 30) for making the pool to identify any effect of using different numbers in the pool

on the genotype-phenotype association study. An equal number of individuals of each sex were used to make the pools in each family. When pools were constructed with 30 individuals, some families showed very strong association of the SNPs to the phenotypic sex (Families 3, 4, 13), some showed weaker association (Families 2, 6) and some showed no association (Families 5, 19). In case of pooled samples where 15 individuals were used (three families), one family showed strong association (Family 7), one family showed weaker association (Family 10) and one family did not show any association (Family 14; **Table 4.1**; **Appendix Figure C4.1**). The polymorphic loci were distributed across the genome in all of the 19 GIFT families, with all 22 linkage groups fairly evenly represented. This suggests that the variation in the number of individuals per pool had little effect in the present genotype-phenotype association study.

Other potential sources of variation can come from sequencing errors or the use of restriction digestion-based NGS studies where allelic dropout occurs if the restriction site contains polymorphism (Gautier *et al*. 2013a). It has been noted for standard RADseq that increasing the sample size in a pool increases the occurrence of allelic dropout by increasing the chances of mutations within the restriction site. This type of problem is yet to be explored for ddRADseq where two restriction enzymes are used; it is more likely to increase the probability of allelic dropout if the restriction enzyme cut sites are polymorphic. In this family-based association study using pool-based ddRADseq, c. 80 % of the generated loci were filtered out (those that did not follow the Mendelian segregation, non-polymorphic, multiple SNPs per locus and more than two alleles per SNP); and the remaining c. 20 % of RAD loci were used and showed clear association with phenotypic sex in Stirling and GIFT families. In the Stirling population > 13 K to 15 K RAD loci and > 1.4 K to 2.5 K bi-allelic polymorphic loci were generated from the pool-based ddRADseq analysis. Palaiokostas *et al*. (2015) also found mostly similar numbers of ddRAD loci

(10,303) and bi-allelic (loci with two alleles) polymorphic loci (1,279) for genotype-phenotype association using individual-based ddRADseq from the same Stirling stock.

It has also been noted that the power of the pooling strategy can be improved by making up multiple pools from the same individuals, replicating the pools for genotyping or sequencing, or increasing the sequencing read depth (Sham *et al*. 2002; Robasky *et al*. 2014). In the current experiment, pooled samples were replicated 4 times in the first run and no replication was used in the second run. More reads were obtained in the first run but the replication did not show any significant effect on the number of polymorphic filtered loci retrieved for further analysis (**Table 4.4**). The replication also did not show any effect on the strength of association between the phenotypic sex and SNPs in the GIFT families studied.

From the above discussion it is suggested that the pre-extraction pooling of tissue samples proved to be an efficient method in discovering markers linked to the phenotypic sex using the ddRADseq technique.

4.5.2 Identification and verification of the sex-determining region in GIFT

This is the first genomic study about the sex determination or sex-linked markers identification in the GIFT stock so far. In the present study, the extension of BSA to ddRADseq was used to sequence, locate and identify a major sex-determining locus in GIFT. LG23 showed a significant association with the phenotypic sex in the majority of the GIFT families.

GIFT was developed through selective breeding from mixed wild (Egypt, Ghana, Kenya and Senegal) and domesticated (Israel, Singapore, Taiwan and Thailand) populations of Nile tilapia. It has been noted that before GIFT establishment, the domesticated populations

of *O. niloticus* in Asia were thought to have lost purity through gene introgression from *O. mossambicus* (Macaranas *et al*. 1986).

Six families were chosen based on the strength of the association of the markers to LG23 from BSA-ddRADseq analysis. Five tightly sex-linked SNP markers from LG23 were analysed which showed the highest association with phenotypic sex where the sire was informative. In the family-based association study, the *ss2017360173* marker showed significant association with the phenotypic sex where only the sire was informative and there were only three departures from the expected phenotypic sex were observed out of 190 offspring. The markers *ss2017360134* and *ss2017360178* were also strongly linked to the phenotypic sex where the sire was heterozygote and four departures were observed (160 in total). In the population-based study, males were heterozygous and females were homozygous for the marker *ss2017360173* except six males and four females.

Eshel *et al*. (2011, 2012) identified two sex-linked microsatellite markers (*UNH898* and *ARO172*), which are located 3.2 cM apart in LG23, flanking the *Amh* gene. Alleles 267 and 274 for *UNH898* and *ARO172* markers respectively were strongly associated with the male phenotype in all the GIFT samples analysed. Insertions or deletions or SNP in the variation of *Amh* gene (Li *et al*. 2015) showed linkage to the male sex determination in the GIFT individuals. The insertions and deletions found in one variant of the *Amh* gene, i.e. *Amh∆y,* were same as described by Li *et al*. (2015) in the Japanese strain of Nile tilapia, and were found to be linked to male sex determination in GIFT. Differences were found in case of the other duplicated copy *Amhy* in GIFT. GIFT individuals showed Y-linked promoter deletion, a Y-linked 3 bp insertion in exon 0 (not mentioned in either Eshel *et al*. 2014 or Li *et al*. 2015) and the lack of a single base change in exon II. The SNP in exon II was thought to have critical role in the male sex determination in Nile tilapia derived from Japanese strain

(Li *et al*. 2015). In case of the families that did not show any association (family 19) or weaker association (family 1, 2; **Appendix Figure C4.1**) in BSA-ddRADseq analysis, a significant association was found between phenotypic sex and all the markers in LG23 in the later analysis of individual samples for each family.

Some of the departures of markers from the expected phenotypic sex were common for all the markers tested (**Appendix Table C4.6**). From the individual analysis of six GIFT families, it was found that those families showed weakest or no association in BSA-ddRADseq analysis where more phenotypic males were found with the female genotype (**Table 4.7**). In case of other families which also have weaker or no association with LG23 markers, this could also be because of the occurrence of recombination within the family, human error in the assessment of phenotypic sex (for example four phenotypic males – based on microscopic sexing - had female genotypes and based on visual external sexing those individuals were females; the same thing happened for another four phenotypic females) or because of the variation of representation of the genetic material from each tissue. Those factors alone or combined with other factors such as minor genetic or environmental factors, or evidence of variation in sex determination among the families in Nile tilapia (GIFT tilapia was developed from multiple stocks of Nile tilapia and the base population of GIFT was also reported to be introgressed) could influence the weaker or absent genotype-phenotype association from the BSA-ddRADseq analysis in some GIFT families.

Temperature dependent sex-ratio is also evident in Nile tilapia and loci in LG1, 3, 20 and 23 showed polymorphism that are found to be associated with temperature effects on sex-ratio (Lühmann *et al*. 2012; Wessels *et al*. 2014; Palaiokostas *et al*. 2015). Wessels *et al*. (2014) identified an allelic variant (*ss831884014*) in exon VI of *Amh* gene (LG23) which was

found to be responsible for autosomal and temperature-dependant sex-reversal in their Stirling-derived Nile tilapia population. This SNP (G > C) arises from a missense mutations leading to an amino acid change from glutamine to glutamic acid. The homozygous CC genotype significantly increased the male proportion in high temperature-treated groups. GIFT and Stirling broodstock were genotyped for this variant and there was no allelic segregation observed between the males and females. Following sequencing of six individuals, no polymorphism was observed in that position.

The different sex-determining loci in different strains of Nile tilapia indicates that different strains possess different sex chromosomes or different sex-determining genes might be responsible for male sex determination in different populations of Nile tilapia. It has been recorded that the three different strong candidate genes responsible for male sex determination in three closely related species of Medaka (Matsuda *et al.* 2002; Myosho *et al.* 2012; Takehana *et al.* 2014). GIFT was surprisingly found to be uniform in sex determination, given its complex genetic structure; and a signle *Amh* gene (LG23) was found as a candidate gene responsible for male sex determination across the population. There was no evidence of harboring *Amh* gene in LG1 which suggest that the "jumping" gene theory might not imply in Nile tilapia sex determination. On the other hand, it has been recorded in salmonids that during evolution, genes have jumped into different autosomes resulting in new Y chromosomes (Yano *et al*. 2013). It has been recorded that the *Sdy*, a master sex-determining gene in rainbow trout has been transposed by movable elements and was found to be conserved in all salmonids (Yano *et al*. 2012, 2013).

The two microsatellite markers (*UNH898*, *ARO172*) and a SNP marker (*ss2017360173*) flanking the *Amh* genes were found as tightly sex-linked markers and were able to discern YY GIFT from XY. This is the first study, provided the tightly sex-linked markers for the

GIFT with very strong evidence. Their physical position is in a range of c. 804 K bases in the published Nile tilapia genome. This genomic region had 32 identified genes, 14 of them annotated and the rest not annotated. It has been reported that the members of the transforming growth factor beta (TGF-β) signalling pathway (*Gsdfy*, *Amhy* and *AmhrII*) could be part of a common pathway for sex determination in most of the fish with exceptions for some species such as salmonids where a single gene (*Sdy*) was found to be conserved across the whole group (Hattori *et al.* 2012; Kamiya *et al.* 2012; Myosho *et al.* 2012). Variations of the *Amh* gene (either *AmhΔy* or *Amhy*), a member of the TGF-β superfamily, could also be the possible candidate gene for male sex determination in GIFT.

4.6 Conclusions

Analysing sex determination in one or a few families, as done in most previous RADseq or ddRADseq based sex determination studies, was not considered to be sufficient in analyzing the sex determination in GIFT given the complex origin of the GIFT population (base stock developed from multiple populations) and different major sex-determining loci in other Nile tilapia populations. The strategy was to use BSA to cover many GIFT families and to obtain information on sex determination at the population level.

Pre-extraction pooling of tissue samples for BSA-ddRADseq proved to be an efficient method as an alternative to individual sequencing or post-extraction pooling in family-based association studies. This allowed screening of multiple families in the GIFT population, leading to mapping of a single sex-determining locus and sex-linked SNP markers with reduced experimental costs. This method could be used to map a range of other loci affecting important phenotypic traits using different NGS platforms.

This is the first genomic study of sex determination in the GIFT tilapia population and only one locus (LG23) was identified as the major sex-determining locus in GIFT across the population. A set of tightly sex-linked SNPs were identified and two previously identified microsatellites and markers in the variation of *Amh* gene were suggested as sex-linked markers for the GIFT population, which would allow marker-assisted selection in GIFT for the first time to produce all-male population for controlling sex-ratio in culture systems. No efforts were made to address the question of whether one of the Y-linked *Amh* variants in GIFT is actually the sex-determining gene, but the missense SNP in exon II of *Amhy*, proposed by Li *et al*. (2015) to be key in male determination, was absent in the GIFT individuals analysed in the present study.

Acknowledgements

The authors acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance (CSC reference BDCA-2013-5). The work was also partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions. We are grateful to Kerry Bartie for technical support.

Chapter 5

LG23-linked sex determination in Nile tilapia

Sex determination varies within the Stirling population of Nile tilapia (*Oreochromis niloticus*)

Taslima K.[1,2], Khan M.G.Q.[2], McAndrew B.J.[1], Penman D.J.[1]

[1]Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2]Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh

Status: To be submitted

Contributions:

Experimental design, experiment set up, tagging of fish as future broodstock and mucus sampling, crossing and fish rearing, phenotypic sex identification, sample collection, DNA extraction, genotypic sex identification through SNP genotyping system (KASP) and standard PCR, data compiling, association analysis were conducted by the author of this thesis. The first draft of the manuscript was written by the author of this thesis who was also entirely involved in the following corrections of the manuscript. The other co-authors contributed in the experimental design, fish rearing, phenotypic and genotypic sex identification and editing the manuscript.

5.1 Abstract

Tilapias (family Cichlidae) are the second most important group of finfish in global aquaculture by production volume. The Nile tilapia (*Oreochromis niloticus*) accounts for the majority of this. This species has been the subject of much research on sex determination due to the problems caused by maturation and reproduction in culture before harvest, and because of the complexity of the sex determination system(s). The emergence of linkage mapping and genomics has allowed detailed analyses, leading to evidence of more than one sex-determining locus and variation between populations. The earliest results indicated an XX/XY locus in linkage group (LG) 1 in a population held at the University of Stirling, derived directly from Lake Manzala in Egypt. More recent studies on the same population detected a locus in LG20 that affected sex-ratio in some families, and was associated with the effects of elevated temperature on sex-ratio. A male-specific variant of the *Amh* gene in LG23 has been associated with sex in a population held at the Agricultural Research Organization in Israel (stated to be derived from the University of Swansea, which was in turn derived from the Stirling population). A screening of the Stirling Nile tilapia population for this male-specific *Amh* variant (Y-linked deletion in *Amh* exon VII) showed that this was associated with male phenotypic sex and present at low frequency. Four Stirling Nile tilapia families were produced, one from XX♀ × XX♂ (family 1) and two from XX♀ × YY♂ (family 2 and 3) and one from XY♀ × YY♂ (family 4) crosses based on the LG1 marker, and sire for all four families was also informative for LG23 marker; and progenies were tested for LG1 (*Oni23063*), LG20 (*Oni3161*) and LG23 markers (Y-linked deletion in *Amh* exon VII). All the progeny from family 1 were phenotypically male (except 2) and had Y-linked deletion which suggest that the phenotypic sex of the progeny in this family was linked to the LG23 (non-informative for LG1). On the other hand in family 2 and 3, all progenies were phenotypically male and were found to be linked to the LG1

marker, no influence from LG23 marker. In family 4, two loci (LG1 and LG23) were found to be influencing the male phenotype development in this family. On the other hand LG20 was not found to show association with sex in any of the families studied. It thus appears that the LG23 sex-determining locus is active in this population, although the male-determining haplotype is rare, and that this may have led to some confusion in earlier attempts to develop YY males. Increased knowledge of sex determination in this species is expected to aid marker-assisted selection (MAS) to produce all male fish to prevent unwanted reproduction in tilapia aquaculture.

Keywords: Nile tilapia (*Oreochromis niloticus*), *Amh* gene, sex determination, sex control, aquaculture

5.2 Introduction

Tilapias are the second largest contributors to the world finfish aquaculture production after carps. Nile tilapia (*Oreochromis niloticus*) is the most widely farmed species among the many tilapia species cultured. Monosex male tilapia production is highly desirable not only for increasing the production (males grow faster than females) but also to avoid overcrowding problems due to unwanted reproduction in mixed-sex stocks in many culture systems. Production of all male tilapia was achieved through manual separation of males from females by visual observation of genital papilla (Popma *et al*. 1984), interspecific hybridization (Hickling 1960; Fishelson 1966; Hulata *et al*. 1983), hormonal sex-reversal (Mair *et al*. 1995, 1997; Beardmore *et al*. 2001) or genetically all male tilapia can be produced by crossing XX female with YY males (Mair *et al*. 1997). Sex-linked or sex-specific molecular markers need to be identified to aid the development of reliable genetically all-male tilapia.

In fish, the largest group of vertebrates, sex-determining mechanisms are not always conserved and have changed frequently during evolution (Böhne *et al*. 2013). Closely related fish species or different populations of the same species can have different sex-determining genes (Mank *et al*. 2006). So far six master sex-determining genes have been identified in fish: three different genes in three different species of medaka, *Dmy*/*Dmrt1by* in *Oryzias latipes* and *O. curvinotus* (Matsuda *et al*. 2002, 2003; Nanda *et al*. 2002), *Gsdfy* in *O. luzonensis* (Myosho *et al*. 2012), *Sox3y* in *Oryzias dancena* (Takehana *et al*. 2014), *Sdy* in the rainbow trout (*Oncorhynchus mykiss*) and other salmonids (Yano *et al*. 2012, 2013), *Amhy* in the Patagonian pejerrey (*Odontesthes hatcheri*) and in Nile tilapia (Hattori *et al*. 2012; Eshel *et al*. 2014; Li *et al*. 2015) and the same *Amh* receptor type II (*Amhr2*) has been identified in three different *Takifugu* species (*T. rubripes*, *T. pardalis* and *T. poecilonotus*) (Kamiya *et al*. 2012).

Anti-Müllerian hormone (*Amh*), a member of the transforming growth factor beta (TGF-β) superfamily, is secreted by the Sertoli cells and is responsible for the regression of Müllerian ducts during testis development in mammals, birds, and reptiles (Josso *et al*. 2001; Teixeira *et al*. 2001; Rey *et al*. 2003). Fish do not have Müllerian ducts but they have *Amh*, which is considered to be autosomal in some species (Miura *et al*. 2002; Yoshinaga *et al*. 2004; Wu *et al*. 2010; Hattori *et al*. 2012). Teleost *Amh* has seven exons (Halm *et al*. 2007) whereas there are five in mammals (Cate *et al*. 1986) and birds (Eusèbe *et al*. 1996).

Like most fish species, the tilapia karyotype does not show any differentiation between the sex chromosomes (Devlin and Nagahama 2002). Tilapia was thought to have complex sex-determining systems and both XX/XY and WZ/ZZ sex-determining systems have been observed in tilapias. Different sex-determining regions have been mapped to LG1 (XY), LG3 (XY and WZ) and LG23 (XY) (Lee *et al*. 2003, 2004, 2005, 2011; Cnaani *et al*. 2008;

Eshel *et al*. 2012). Sex in Nile tilapia is mainly controlled by the major genetic factors (XY, Mair *et al*. 1991) and minor autosomal factors can also override the genetic factors (Müller-Belecke and Hörstgen-Schwark 1995). Environmental factors (mainly high temperature) may also play a significant role in changing the phenotypic sex of female to male (Baroiller *et al*. 2009a).

Sex determination analysis has been conducted in Stirling families of Nile tilapia derived from Lake Manzala, Egypt with balanced sex-ratio, using a genotyping by sequencing technique, restriction-site associated DNA (RAD) sequencing, to screen SNPs linked to sex (Palaiokostas *et al*. 2013a). They identified a major sex-determining region in LG1 and two SNP markers (*Oni23063* and *Oni28137*) have been shown to tightly link to the phenotypic sex. Deviations from the expected sex-ratio were been observed in normal crosses (XX and XY) and it could be because of the hormone/temperature effects, complex genetic sex determination or interaction between genetic and environmental factors (Penman and Piferrer 2008). Another study was conducted for a few families of the same population of Nile tilapia (Stirling) with skewed sex-ratio at 28^{o}C and 36^{o}C, using a variant of RADseq, double-digest RADseq (ddRADseq). A sex QTL on a different chromosome (LG20) was identified which was found to be responsible for the skewed sex-ratio in these families (Palaiokostas *et al*. 2015).

Sex-linked microsatellite markers have been mapped to three different linkage groups such as LG1, 3 and 23 in different temperature treated families of Nile tilapia in the University of Göttingen (Germany), which was derived from the University of Stirling population (Lühmann *et al*. 2012). In the same population, an allelic variant in *Amh* exon VI (LG23) was identified as a sex QTL in a line selected for thermosensitivity (Wessels *et al*. 2014). Other studies have identified several sex-linked markers in the same species, which are very

close to *Amh* gene in LG23 (Eshel *et al*. 2011; University of Swansea population). A male-specific duplicated copy of *Amh* (termed *Amhy*) was identified in an Israeli population of Nile tilapia which was derived from the University of Swansea (which in turn was derived from the University of Stirling; Eshel *et al*. 2014). A 233 bp deletions in exon VII was found to be linked to the male sex determination. Li *et al*. (2015) identified the two variants of *Amh* gene on the Y chromosome, which was found to have significant role in male sex determination in a Japanese population of Nile tilapia (of Egyptian origin). *AmhΔy*, a duplicated copy of *Amh* gene (termed *Amhy* by Eshel *et al*. 2014); several insertions and deletions were identified in the promoter region and a 5 bp insertion in exon VI of *AmhΔy*. Li *et al.* (2015) also identified another tandem duplicate *Amh* gene variant (termed *Amhy*) which is located immediately downstream of *AmhΔy* on the Y chromosome. *Amhy* does not have promoter sequences (5608 bp deletion) and the exon sequence of X-linked *Amh* is identical to the Y-linked *Amhy*, except for a missense SNP (C/T) in exon II in *Amhy*, which changes an amino acid (Ser/Leu92). This SNP (*Amhy*) was proposed by Li *et al*. (2015) to have a critical role in male sex determination in the Japanese population of Nile tilapia. Sex-linked SCAR (Sequence Characterized Amplified Region) markers close to *Amh* gene in LG23 were also demonstrated in the same population (Japan) derived from Egypt (Sun *et al*. 2014). Sexually dimorphic expression of *Amh* in brain and gonads was also reported during the early developmental stage (10 dpf) in Nile tilapia (Poonlaphdecha *et al*. 2011).

Given the emerging evidence of different sex-determining loci in Nile tilapia, including in populations derived from the Stirling, this study was conducted with the Stirling population of Nile tilapia by screening of broodstock first (both red and wild type) using tightly sex-linked markers in LG1, LG20 and LG23 followed by producing progenies from crosses XX♀ × XX♂, XX♀ × YY♂ and XY♀ × YY♂ (genotypes based on LG1 marker, sires for

all these families also had Y-linked 233 bp deletions from *Amh* exon VII in LG23) to test the segregation pattern of the sex-linked markers to their progenies.

5.3 Materials and methods

5.3.1 Sample collection from broodstock and genomic DNA extraction

A range of male and female broodstock from wild-type and red Nile tilapia (*Oreochromis niloticus*, University of Stirling) were selected to screen the genotypic sex of fish using the sex-linked markers in LG1, LG20 (previously found to be linked to the sex in the same population) and LG23 (not assigned before in Stirling population). Stirling Nile tilapia was brought from a population from Lake Manzala, Egypt in 1979 (31°16′N, 32°12′E). Randomly selected 18 and 15 wild type males and females; 17 and 12 red type males and females respectively, and 2 clonal red females were used to test for LG1, LG20 and LG23 markers. For screening more broodstock for LG23 marker, 16 wild type (male – 8, female - 8) and 16 red type (male – 8, female - 8) broodstock were also tested. Minimally invasive DNA samples (skin mucus) were collected from each fish using filter paper as described in Chapter 3 and preserved in 100 % ethanol. Genomic DNA was extracted using the salt precipitation method described in Chapter 3. In short, filter paper soaked with mucus was digested overnight (at 55^{o}C) in lysis solution (SSTNE/SDS) and proteinase K (10 mg/mL). Digested samples were treated with RNaseA (2 mg/mL) followed by protein precipitation using 5 mM NaCl. Isopropanol precipitated DNA was dissolved in 5 mM Tris. Quality and quantity of genomic DNA was assessed by Nanodrop (ND-1000) spectrophotometer (Labtech International Ltd, UK) and 1 % agarose gel electrophoresis.

5.3.1.1 Genotyping of broodstock using SNP markers

Broodstock were then tested with tightly sex-linked SNP markers in LG1 (*Oni23063*, Palaiokostas *et al*. 2013a) and LG20 (*Oni3161*, Palaiokostas *et al*. 2015). SNPs were analysed using the fluorescence-based Kompetitive Allele Specific end point-PCR (KASP) genotyping system. SNP assays (based on the provided sequences) and master mix were supplied by LGC Genomics Ltd (UK). Primer sequences are mentioned in **Appendix Table C4.4**. Briefly, 5 µL PCR reaction volume was prepared with ~ 25 ng genomic DNA and the PCR conditions were the initial activation at 94^{o}C for 15 min, 10 cycles of denaturation at 94^{o}C for 20 sec, annealing and extension at 65^{o}C to 57^{o}C for 1 min (dropping 0.8^{o}C per cycle) and finally 34 cycles of amplification at 57^{o}C for 1 min. Following PCR, fluorescence-labelled PCR products were measured on a Techne Quantica® machine (Barloworld Scientific Ltd UK) and data was analysed using inbuilt Quansoft software (version 1.1.21). The genotype of each individual was identified based on the fluorescence signal (bound to the respective allele during PCR) detected by the machine.

5.3.1.2 Genotyping of broodstock using *Amh* exon VII deletion in LG23

Broodstock were also analysed for the *Amh* exon VII deletion (Y-linked in LG23) using the standard PCR protocol (primer sequences are mention in **Appendix Table C4.5**). PCR was carried in a 5 µL reaction volume (~ 25 ng genomic DNA) and the cyclic conditions were the initial denaturation at 95^{o}C for 1 min, 35 cycles of denaturation at 95^{o}C for 15 sec, annealing at 60^{o}C for 15 sec and extension at 72^{o}C for 30 sec. Following PCR amplified products (3 µL) were assessed using 1.5 % agarose gel electrophoresis.

5.3.2 Production of families and sex identification

Four families were produced, one from XX♀ × XX♂ and two from XX♀ × YY♂ and one from XY♀ × YY♂ crosses based on the LG1 marker. **Table 5.1** shows the parental genotypes for LG1 (*Oni23063*), LG20 (*Oni3161*) and LG23 (*Amh* exon VII deletion); and some were missing (because they were not alive). To produce each family, mature females and males were kept in individual tanks and were fed with commercially available trout feed (Skretting, Preston, UK). Once the female was found to be ready to spawn, ovulated eggs were stripped manually from the female and fertilised *in vitro* by the milt (contained sperm) from the male. Fertilised eggs were incubated in downwelling incubators until the larvae had absorbed the yolk sac. Fry from each family were transferred to the tanks in recirculatory systems, fed *ad libitum* and reared for three months, when phenotypic sex was identified microscopically (30 fish from the first three crosses and 43 fish from the last cross, **Table 5.1**) using acetocarmine gonad squash method described in Guerrero III and Shelton (1974). A fin clip from each of the progeny was fixed to 100 % ethanol for sex-linked marker study.

Genomic DNA was extracted using the salt precipitation method and the DNA integrity was checked following the protocol described above. Each sample was analysed for sex-linked SNP markers in LG1 and LG20 using KASP genotyping system, and the marker in LG23 using the PCR protocol described above.

Table 5.1 Families tested, including genotype based on LG1 (sex linked SNP, *Oni23063*), LG20 (sex-linked SNP, *Oni3161*) and LG23 markers (Y-linked deletion), and the number of progeny tested.

Family	Genotypes (LG1)	Genotypes (LG20)	Genotypes (LG23)	Number of progeny
1	G/G (XX)♀ × G/G (XX)♂	C/C♀ × C/T♂	Y-linked deletion in sire	30
2	G/G (XX)♀ × A/A (YY)♂	C/C♀ × T/T♂	Y-linked deletion in sire	30
3	G/G (XX)♀ × A/A (YY)♂	?	?	30
4	A/G (XY)♀ × A/A (YY)♂	T/T♀ × T/T♂	Y-linked deletion in sire	43

5.4 Results

5.4.1 Broodstock screening for sex-linked markers

All the females were homozygous (G/G) for the LG1 marker, whereas the majority of the males were heterozygous (A/G) for this marker, regardless of colour phenotypes (**Appendix Table C5.1**). Nine of the males were homozygous for the LG1 marker (G/G, putative sex-reversed male: one of them were progeny tested and approximately 95 % of the progenies were female). Two of the males were homozygous for the LG1 Y-linked allele (A/A, **Appendix Table C5.1**), and progeny information was available for these males and 100 % phenotypic males were found when crossed with XX female for LG1 (concluded to be YY supermale for LG1). The broodstock LG1 genotypic data suggested that the LG1 marker was, as expected to be a major sex-determining locus, linked to the phenotypic sex in the Stirling population. Association analysis (Fisher's exact test, **Table 5.2**) suggested that a significant association was observed in terms of the allelic segregation for the marker in

LG20 (*Oni3161*) for the broodstock with the female expected genotype for LG1 (P = 0.0061), while for the broodstock with the male expected genotype was not significant (P > 0.999). The results from the LG23 marker study showed that the Y-linked *Amh* deletion was not observed in females, as expected, and all females were homozygous for the X-linked *Amh* gene irrespective of body colour (**Figure 5.1**). On the other hand 3 XX males, 2 XY males and 2 YY males had the LG23 Y-linked *Amh* deletion (which was only found in these phenotypic males), suggesting that the LG23 male (Y) sex-determining gene is present in the Stirling population, but at a low frequency (**Figure 5.1**). The two YY male for LG23 (also YY for LG1) produced 100 % phenotypic male progenies when crossed with XX female (based on the LG1 marker and 100 % phenotypic males were found from the progeny test).

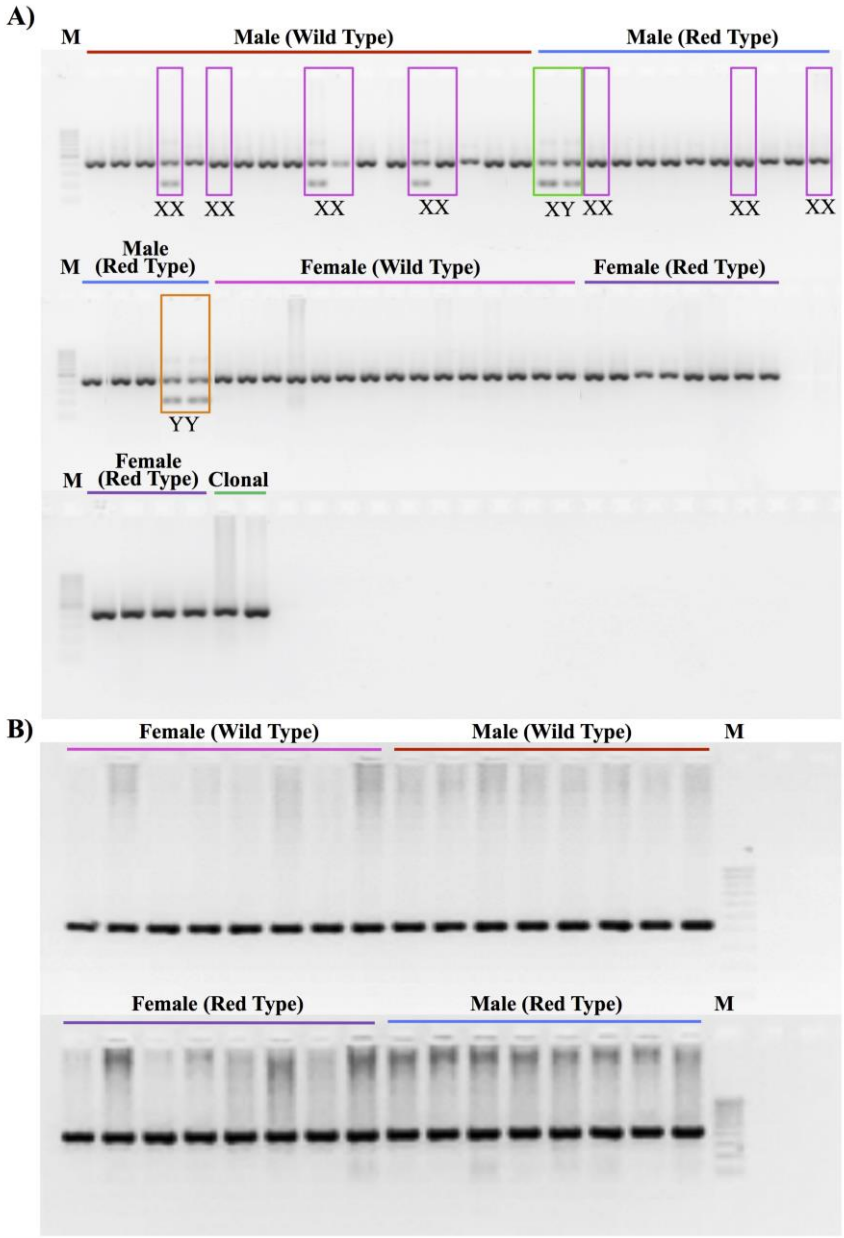


Figure 5.1 Amplified PCR products on 1.5 % agarose gel using *Amh* exon VII marker (LG23) to detect the Y-linked deletion of 233 bp for multiple wild and red of male and female Nile tilapia. A) The X-linked 439 bp band was evident in all the males and females irrespective of the wild/red types. Individuals within purple coloured boxes are XX males based on the LG1 marker. Three of them had the LG23 Y-linked 233 bp deletion. Two LG1 XY males (green colour box) and two LG1 YY males (yellow colour box) also had this Y-linked deletion. B) 16 more males and 16 more females from wild and red types of Nile tilapia were also analysed. All showed the 439 bp band (*Amh*) only. M- 100 bp molecular ladder.

Table 5.2 Allelic combinations between SNP markers on LG1 (*Oni23063*) and LG20 (*Oni3161*) for the broodstock analysed with the statistical test.

	LG1: G/G		**Fisher's**
	Female expected genotype		**exact test**
	Male	Female	
LG20: C/C	0	7	
LG20: C/T	4	11	$P = 0.0061$
LG20: T/T	5 (3 had Y-linked deletion)	11	
	LG1: A/G		
	Male expected genotype		
	Male	Female	
LG20: C/C	5	0	
LG20: C/T	12	0	$P > 0.999$
LG20: T/T	7 (2 had Y-linked deletion)	0	

5.4.2 Family 1: XX♀ and XX♂ cross (based on LG1)

Twenty-eight progeny were phenotypically male and two were phenotypically female from the XX♀ × XX♂ cross (based on the LG1 marker), although only female progeny were expected. The genotypic sex of all the progeny, based on the LG1 marker, was XX as expected, which suggests that the LG1 sex-determining locus was not influencing phenotypic sex in this family (**Table 5.3**). No association was found between phenotypic sex and the LG20 marker (**Table 5.5**). All of the progeny (including the two phenotypic females) were found to have the LG23 Y-linked 233 bp deletion, suggesting that the sire was homozygous for the LG23 Y haplotype (i.e. an LG23 YY male, **Figure 5.2**).

Table 5.3 Allelic segregation for the SNP marker (*Oni23063*) in LG1 and *Amh* exon VII marker in LG23 for the progenies of the first three families.

		LG1	
		Female expected genotype (G/G)	
	LG23	**Male**	**Female**
Family 1	**No *Amh* deletion**	0	0
	Y-linked deletion	27	2
		Male expected genotype (A/G)	
		Male	**Female**
Family 2	**No *Amh* deletion**	16	0
	Y-linked deletion	14	0
Family 3	**No *Amh* deletion**	17	0
	Y-linked deletion	13	0

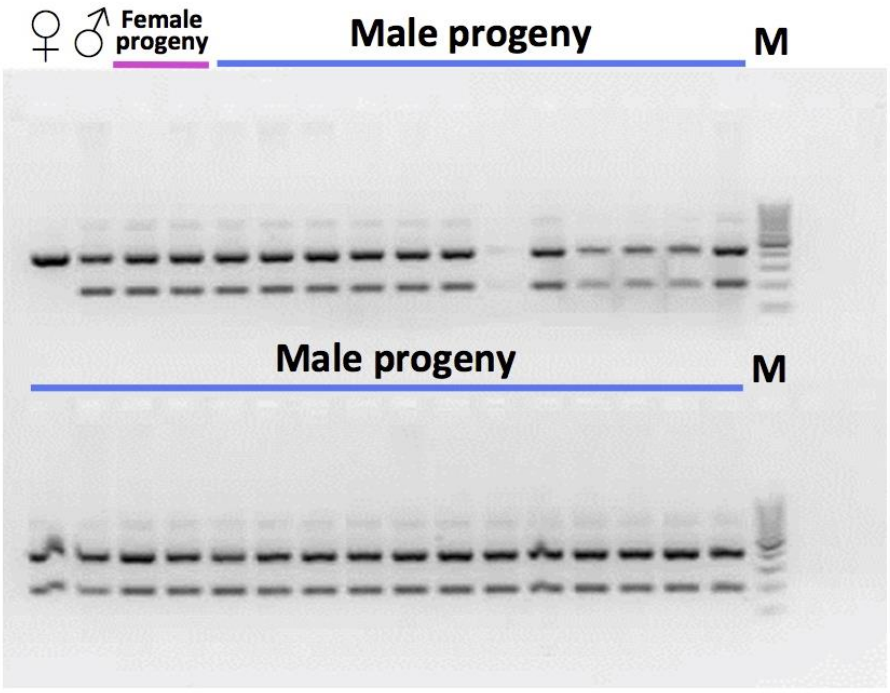


Figure 5.2 Amplified PCR products on 1.5 % agarose gel using *Amh* exon VII marker in LG23 for family 1. ♀ - dam and ♂ - sire. Dam had only X-linked band whereas sire had both X and Y-linked band. All the male progeny had X-linked 439 bp band and Y-linked 233 bp deletion (one did not amplify). Two female progeny also had the same pattern. M - 100 bp molecular ladder.

5.4.3 Families 2 and 3 (see Table 5.1): XX♀ and YY♂ cross (based on LG1)

Progeny from families 2 and 3 (XX♀ × YY♂ from LG1 genotypes) were phenotypically male as expected based on LG1 (heterozygous for the LG1 SNP marker; **Table 5.3**). At the LG20 marker, in case of family 2 all the progeny were heterozygous (C/T) while the parents were homozygous for each allele (dam – C/C and sire – T/T; **Table 5.1**, **5.5**). Likewise, in family 3 no association was found between sex and the LG20 marker (**Table 5.5**). For LG23 marker, approximately half of the progeny in each of families 2 and 3 (**Figure 5.3**) had the Y-linked deletion (233 bp), suggesting that the sire was heterozygous in family 2 and that one parent was heterozygous in family 3 (presumably the sire: no parental data

available). No association between the LG23 Y-linked marker and phenotypic sex could be detected, as the sire homozygosity for the LG1 Y allele appeared to have been enough to ensure that all progeny were phenotypically male.

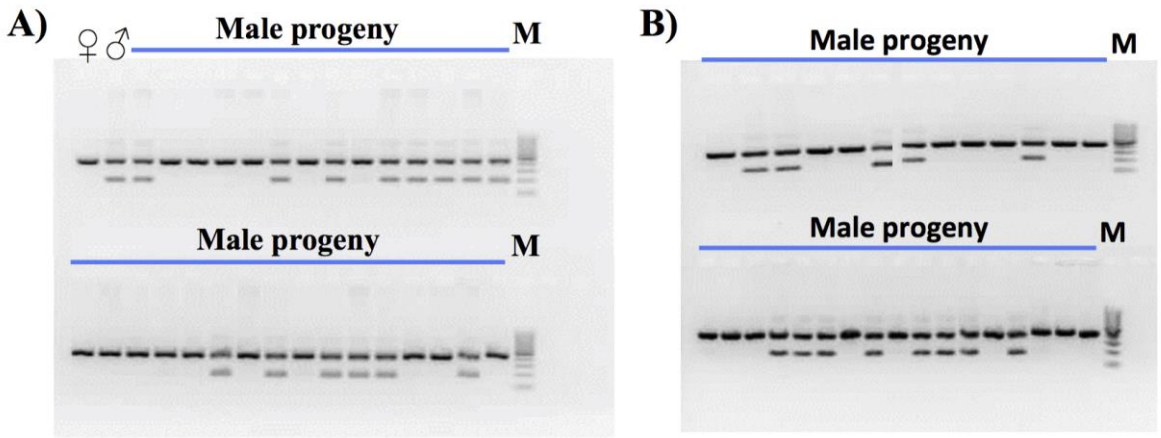


Figure 5.3 PCR products using *Amh* exon VII marker in LG23 visualized on 1.5 % agarose gel. A) Family 2. ♀ - dam and ♂ - sire, 14 out of 30 male progeny showed Y-linked 233 bp deletion. B) Family 3. 13 out of 30 male progeny showed Y-linked 233 bp deletion. Parental samples were unavailable for this family. M- 100 bp molecular ladder.

5.4.4 Family 4: XY♀ and YY♂ cross (based on LG1)

All the progenies from this cross (XY♀ × YY♂ according to the LG1 marker) were phenotypically male. The expected ratio would be 50 % XY males : 50 % YY males, but LG1 marker analysis indicated that 14 were XY males (heterozygous genotype) and the rest (28) were YY males (**Table 5.4**). All of the progeny, like their parents, were homozygous for the T allele of the LG20 SNP marker, so no influence of LG20 was expected on phenotypic sex (**Table 5.1, 5.5**). All the progeny had the 233 bp deletion on the Y chromosome (**Figure 5.4**). The same sire as used for family 2 was used to produce this family. According to the genotype for the LG23 marker and the pattern in the progeny in

Family 2, the sire was heterozygous (i.e. LG23 XY), while the dam was concluded to be homozygous (i.e. LG23 YY neo-female) based on the progeny genotypes.

Table 5.4 Allelic combinations for the *Amh* exon VII marker in LG23 and SNP marker (*Oni23063*) in LG1 for the progeny of family 4.

		LG23	
		Male expected genotype (Y-linked deletion)	
Family 4	**LG1**	**Male**	**Female**
	A/G	14	0
	A/A	28	0

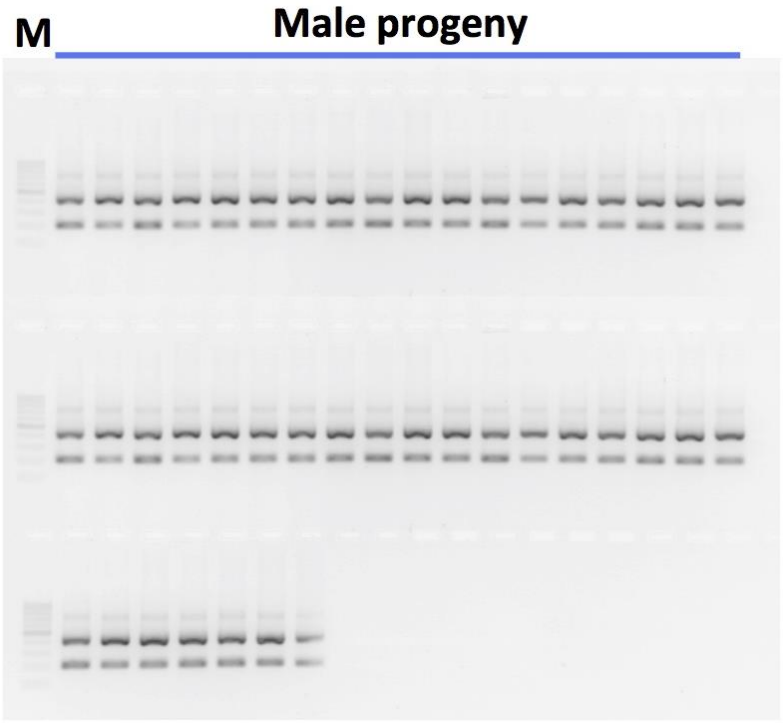


Figure 5.4 Amplified PCR products on 1.5 % agarose gel using *Amh* exon VII marker in LG23 for family 4. All male progeny had both 439 and 233 bp band. M- 100 bp molecular ladder.

Table 5.5 Allelic inheritance for the SNP marker (*Oni3161*) in LG20 to the progenies (phenotypic male and female) of the four families.

	Family 1		**Family 2**		**Family 3**		**Family 4**	
Genotype	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**	**Male**	**Female**
C/C	7	1	0	0	0	0	0	0
C/T	17	1	30	0	15	0	0	0
T/T	0	0	0	0	15	0	43	0

5.5 Discussion

Nile tilapia sex determination is of great research interest due the complexity that has emerged from ongoing study, and also due to the need to control reproduction in many production systems. It has been observed that sex determination in Nile tilapia varies among and within populations. Two major (XY) sex-determining loci, in LG1 (University of Stirling) and LG23 (ARO Israel, originating from University of Stirling) have been identified (Palaiokostas *et al*. 2013a; Eshel *et al*. 2014). Given the evidence of two sex QTLs in LG1 and LG20 in the Stirling population, and a different QTL in LG23 in the Israeli populations of Nile tilapia, which originated from the University of Stirling, this study was carried out to screen the Stirling Nile tilapia population to check if the LG23 male-determining allele was present and functional.

The sex-linked marker in LG1 (*Oni23063*) was found to be strongly associated with the phenotypic sex of the Nile tilapia broodstock studied, as expected. All phenotypic females and most phenotypic males were homozygous (for the X-linked allele) and heterozygous respectively for this marker, but there were a few exceptions among the phenotypic males. Earlier progeny sex ratios from one XX (LG1) and two YY (LG1) males suggested that those males were sex-reversed XX neo-male and YY supermale respectively. No phenotypic females showed Y-linked deletion (*Amh* exon VII, LG23), but three out of nine LG1 XX males, two LG1 XY and two LG1 YY males were found to have this deletion, which suggests that it is present at very low frequency in the Stirling population, and possibly functional (based on the three LG1 XX neo-males carrying this deletion). Five XX neo-males based on LG1 markers remain to be unexplained, which is similar to the findings of Palaiokostas *et al*. (2013a) who also found a low proportion of such fish. Palaiokostas *et al*. (2015) identified another sex QTL in LG20 (linked to the SNP *Oni3161*), that was

associated with masculinization of LG1 XX genetic females (particularly following elevated early temperatures, which was not the case here). LG20 results from the present study suggested that it had significant influence on deviation of phenotypic sex (female to male) from the expected genotype for the broodstock analysed (**Table 5.2**).

To extend this study, four families were produced where the sire had the Y-linked LG23 deletion (parents samples were not available for family 3, **Table 5.1**). In the first family, nearly all of the progeny were male, although only females were expected based on the LG1 marker, which suggested that the LG1 XY locus (Palaiokostas *et al*. 2013a) was not determining the sex in this family. All of the progeny showed the LG23 Y-linked deletion (as well as the longer copy of this exon fragment, which suggested that the sire had the LG23 YY genotype and that this was determining phenotypic sex in this family. The two female progeny in this family also had the LG23 Y-linked deletion. This was not surprising in the sense that Mendelian inheritance would predict this from the sib genotype ratio; it is not uncommon for YY males to produce a low percentage of female progeny (Mair *et al*. 1997; Abucay *et al*. 1999; Sarder *et al*. 1999; Karayücel *et al*. 2004; Shirak *et al*. 2006; Baroiller *et al*. 2009b). The phenotypic sex of the next two families was associated with LG1 (LG1 YY sire; all-male progeny) and the LG23 Y-linked deletion was found in approximately half of these progeny, suggesting that the sire was an LG23 XY in both cases (it was not possible to determine this in family 3, but this was more likely that the dam having this genotype), but this deletion could not be associated with phenotypic sex in these families due to their all male nature.

The progeny in the fourth family were all male. The results showed that the sire was homozygous for the Y-associated SNP allele in LG1 and for the Y-associated *Amh* deletion in LG23, while the dam was an LG1 XY neo-female. Because of this it appears that the all

male nature of the progeny was determined by the effects of both loci – it was not possible to distinguish any separate effects of each locus.

Several sex-linked microsatellite markers in LG23 have been identified which are closely linked to the *Amh* gene in the Nile tilapia population from Israel (Eshel *et al*. 2011, 2012). Later on they identified a deletion (233 bp) in exon VII of *Amh* gene (termed as *Amhy*), which was male-specific in this population (Eshel *et al*. 2014). It would be worth mentioning that two of the sex-linked microsatellite markers (*UNH898* and *ARO172*) in LG23 have male-specific alleles that make distinguishing between LG23 XYs and YYs easier using these microsatellite markers than the *Amh* exon VII deletion. In a Chinese population of Nile tilapia obtained from Prof. Nagahama (Laboratory of Reproductive Biology, National Institute for Basic Biology, Okazaki, Japan), which was introduced from Egypt, SCAR markers in LG23 (close to *Amh*) were found to be associated with phenotypic sex (Sun *et al*. 2014). Another study was conducted with the same population in which two variants of *Amh* gene, termed as *Amh*Δ*y* (*Amhy* by Eshel *et al*. 2014) and *Amhy* (Li *et al*. 2015), were identified. All these populations have the same origin but different sex-determining loci and *Amh* variants have been found in different populations, suggesting that founder/bottleneck effects and/or domestication/selection pressure could be the factors leading to this variation. The Stirling population has retained two XX/XY loci, while the Israeli population derived from it (in two transfers) only appears to have one; The population studied by Li *et al.* (2015) also appears to have been through at least two transfers (there was little evidence of assessment of sex determination across the population in this study). A single sex-linked WZ/ZZ locus in chromosome 4 has been identified in wild Zebrafish populations whereas domesticated population failed to show any major sex-linked locus, with only weaker QTL for sex being identified (Wilson *et al*. 2014). They

postulated that the selection pressure during domestication might have led to the loss of the W allele, the development of a new sex-determining system and/or disclosed pre-existing minor genetic sex-determining loci in Zebrafish.

5.6 Conclusions

In conclusion LG1 was confirmed to be the major sex-determining locus in Stirling Nile tilapia population (LG20 also had effect on phenotypic sex reversal); the distribution of alleles in the population and their segregation patterns in crosses were consistent with the previous findings (Palaiokostas *et al*. 2013a). A sex-linked LG23 Y (male-determining) allele was also evident at very low frequency and showed strong effect in determining sex in one of the Stirling families studied. The outcome of this research emphasises the complexity of sex determination in Nile tilapia.

Acknowledgements

The authors acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance (CSC reference BDCA-2013-5). We are also thankful to Keith Ranson and Kerry Bartie for their technical support in the Tropical Aquarium Facilities and in the Laboratory at the Institute of Aquaculture, University of Stirling.

Chapter 6

Feminisation in Nile tilapia

Estrogen hormone with high temperature induces sex-reversal in Nile tilapia, *Oreochromis niloticus*

Taslima K.[1,2], McAndrew B.J.[1], Penman D.J.[1]

[1]Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2]Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh

Status: To be submitted

Contributions:

Experimental design, hormone feed preparation, experiment set up, rearing of fish, phenotypic sex identification using gonad squash method, sample collection, DNA extraction, genotypic sex identification through SNP genotyping system (KASP) and capillary sequencer (microsatellite), PIT tagging of fish as future broodstock and mucus sampling, DNA extraction from mucus sample and XY neo-female broodstock identification using the same genotypic platform, data compiling, association analysis and general statistics were conducted by the author of this thesis. The manuscript was first drafted by the author of this thesis who was also entirely involved in the subsequent corrections of the manuscript. The other co-authors contributed in the experimental design and editing the manuscript.

6.1 Abstract

Nile tilapia proved to be sensitive to steroid hormone and high temperature during the critical period of gonad differentiation. It seems to be difficult to feminise XY males (inconsistent feminisation rates) than masculinise XX females in Nile tilapia and high temperature feminisation has also been observed in Nile tilapia. Therefore this study aimed to increase the feminisation rates using combined treatment of estrogen hormone (either diethylstilbestrol, DES, or 17α-ethynyloestradiol, EE2) and high temperature (36^{o}C) during the sensitive period of gonad differentiation in Nile tilapia. Mixed sex, XY and XX groups were treated with combined treatment of estrogen hormone and high temperature (36^{o}C) including hormone alone, 36^{o}C and 28^{o}C. The sexual genotype was confirmed using tightly sex-linked single nucleotide polymorphism (SNP) and microsatellite markers in linkage group (LG) 1 and further tested with LG20 marker. Combined treatment of estrogen hormones (either DES or EE2) and high temperature induced significant sex-reversal from male to female than hormone alone in mixed and XY sex progeny. Although combined DES and high temperature showed higher sex-reversal (42 - 46 % and c. 94 % in mixed and XY sex group respectively) than the combined EE2 and high temperature (19 - 37 % and 9 – 89 % in mixed and XY sex group respectively), the survival rate was higher in combined EE2 and high temperature (76 - 98 % and 83 – 85 % in mixed and XY sex group respectively) than to the combined DES and high temperature (24 - 55 % and c. 33 % in mixed and XY sex group respectively). Hormone treatment alone did not induce significant feminisation in the batches produced except one from mixed sex batches and one from XY sex batches. It is worth mentioning that the feminisation rate varied among the batches in the same sex group and among the sex groups throughout the experiment. High temperature did not show any significant influence on masculinisation in mixed sex batches except one (*p*-value 1.90e-02). However significant masculinisation was found in XX sex group with

elevated temperature (*p*-value 1.68e-02). About 15 - 68 % XY neo-females were identified as future broodstock from different batches of treatments in all sex groups using tightly sex-linked markers. This is the first study demonstrating the combined treatment of estrogen hormone and elevated temperature induced feminisation in Nile tilapia and further investigations need to be done on sex differentiation and hormonal regulation to optimize the feminisation protocol.

Keywords: Sex-reversal, estrogen hormone and high temperature, sex-linked markers, sex-ratio, Nile tilapia.

6.2 Introduction

Sex control in aquaculture production can be highly desirable because one sex grows faster than the other and some fish mature at very early age leading to unwanted reproduction. In tilapia males grow 20 % faster and have a lower feed conversion ratio than females. Mixed sex tilapia production in some production systems can result in unwanted reproduction leading to overcrowding problems, reducing the value of the final production (Budd *et al*. 2015). Various strategies have been tested for controlling sex-ratio and maturation, such as exogenous hormones, chromosomal ploidy manipulation, hybridisation, varying environmental and social parameters, and selection of broodstock using a molecular and/or quantitative genetic approach. Exogenous hormones and high temperature have been extensively used in fish to obtain monosex population with desired sex species (with particular sexual genotype) due to its efficiency in changing the phenotypic sex and to elucidate the genetic and endocrine factors involved in the mechanism of sex determination and differentiation (Baroiller and D'Cotta 2001; Piferrer 2001; Baroiller *et al*. 2009a; Singh 2013).

Yamamoto (1969) first mentioned that complete sex-reversal could be obtained through the administration of sex steroids during the critical period of gonad differentiation and thereafter highlighted that the sex steroids should be applied from prior to any sign of gonadal differentiation until after the time when normal sex differentiation occurred. This duration is very species-specific and successful sex-reversal also depends on the dose, type and nature of the hormones used. Since then steroid hormones have been used as a key component to produce monosex male or female populations to control the sex phenotype in species of commercial interest (Singh and Pandey 1995; Devlin and Nagahama 2002; Kazeto *et al*. 2011; Singh *et al*. 2012). The direct hormone treated fish for human consumption (e.g. extensively used in tilapia to produce monosex population) has been restricted in some countries like EU and India. Combination of sex-reversal and genetic manipulation in fish can produce genetically monosex progeny for commercial production. For example hormone treatment can be used as the very first step of eventually producing XX or YY population, which will subsequently produce monosex progeny. XX and YY populations have been successfully developed for tilapia and salmonids (Dunham 2011). The intensity of treatment required for successful feminisation varies among fish families i.e. Salmonidae < Cichlidae < Anguillidae < Belontiidaes < Poecilidae < Cyprinidae; salmonids are easier to feminise whereas cyprinids are more difficult (Piferrer 2001).

Different types of natural/synthetic androgenic and estrogenic hormones have been used to sex reverse to either male or female in different fish species (Piferrer 2001). Estrogen treatment induces male individuals to develop as females; androgen treatments inhibit estrogen synthesis, which in turn triggers testicular development in the female genotype (Guiguen *et al*. 1999). Hormones can be administered in three different ways (dietary treatment, injection or immersion), and the choice of the method sometimes depends on the

economic and commercial practice or the biology of the species (age/size of sexual differentiation, feeding habits).

Tilapia is the second most important group of aquaculture species in the world after carps, and precocious reproduction in mixed sex culture can lead to production losses. Therefore researchers have been trying to elucidate the underlying mechanisms of sex determination and differentiation systems in tilapia. Nile tilapia has a male heterogametic sex determining system (XX/XY) and environmental factors also influence the sex differentiation pathway through sex-reversal where phenotypic sex loses the link to the genotypic sex.

Ovarian differentiation starts earlier than the testis differentiation in tilapia (Nakamura *et al*. 1998; Ijiri *et al*. 2008). An increasing number of germ cells prior to meiotic division is one of the crucial signs of ovarian differentiation, whereas testicular differentiation can be recognized by the onset of efferent ducts. Histologically distinguishable gonads, either into ovary or testes, can be found in tilapia about 20 days after hatching. So the most effective period of inducing masculinisation or feminisation in tilapia using exogenous steroid hormones or high temperature is the first 20 days after the first feeding.

The most common practice (either experimental or commercial purpose) used to produce monosex male populations in Nile tilapia is through dietary administration of sex steroids treatments during the sensitive period of sex differentiation, usually between 10 days post fertilisation (dpf) to 30 dpf (Baroiller *et al*. 2009a). It has been found that the initiation of sex differentiation starts in the brain or in the primordial germ cells before the onset of gonad development (Kobayashi and Iwamatsu 2005; Kobayashi *et al*. 2008; Rougeot *et al*. 2008a, b; Blázquez and Somoza 2010). In Nile tilapia, morphological differentiation of the brain starts from 31 hrs post-fertilisation (hpf) and primordial germ cells can be identified from 46 hpf, suggesting that sex differentiation occurs during embryonic development

(Morrison *et al*. 2001). Therefore immersion techniques with steroid hormones or with elevated temperature have been used in Nile tilapia to induce sex-reversal (Rougeot *et al*. 2008a, b; Gennotte *et al*. 2015). Variable feminisation rates have been found using immersion treatment with hormone in XY Nile tilapia embryos during embryonic development, for example 9.9-26.6 % (at 1 hpf), 48.0-48.6 % (at 24 hpf), and 65.9–65.2 % (at 1+24 hpf) feminisation rates were found for 1000 and 2000 μg EE2/L respectively. On the other hand the sex-reversal rate in XX embryos using masculinising hormone in immersion treatment only reached a maximum of 10 % (Gennotte *et al*. 2015). The use of high temperature treatments during the sex differentiation period induced 50 - 80 % masculinisation in XX Nile tilapia whereas a single treatment of dietary AI or combined treatment of AI and high temperature induced complete masculinisation of XX fry (Kwon *et al*. 2002).

In many fish species it is easier to change the genetic females to phenotypic males than the genetic males to phenotypic females. For example in salmonids all female populations are desirable and inconsistent feminisation rates have been found when treated with estrogen hormones (Simpson *et al*. 1976; Donaldson and Hunter 1982). Changing the sex from genetic male to phenotypic female is also difficult and inconsistent in tilapia (Hopkins *et al*. 1979; Meriwether and Shelton 1981). When the genetic males are treated with estrogen hormones, sometimes ovotestes gonad has been developed in large scale due to incomplete sex-reversal (Jalabert *et al*. 1975; Meriwether and Shelton 1981).

Sex steroids play a significant role in phenotypic sex differentiation through their interaction with androgen or estrogen receptors. In many fish species including tilapia, the aromatase gene (*Cyp19a1a*) is known to have significant role in ovarian differentiation through the production of aromatase enzyme, a key enzyme responsible for the conversion

of androgen to estrogen whereas the *Dmrt1* and *Amh* genes are known to be associated with testicular differentiation in tilapia. Treatments with estrogen during the labile period results in sex-reversal of genetic male to phenotypic female by repressing the *Dmrt1* gene and promoting the expression of *Cyp19a1a* gene (Kobayashi *et al*. 2003, 2008). The 5' flanking region of *Cyp19a1a* gene contains estrogen responsive elements (ERE) and following exposure of estrogen, works directly in the expression of *cyp19a1a* gene through ERE (Yoshiura *et al*. 2003).

Foxl2 is a conserved transcription factor involved in ovarian differentiation in vertebrates and its role in tilapia has been described by Ijiri *et al*. (2008). *Cyp19a1a* is known to be activated by *Foxl2* gene in tilapia and Japanese flounder (Wang *et al*. 2007b; Yamaguchi *et al*. 2007). It has also been known that the estrogen exposure upregulates the expression of *Foxl2* gene thereby upregulating the expression of *Cyp19a1a* (Baron *et al*. 2004; Wang *et al*. 2007b).

High temperature has a strong masculinising effect if tilapia fry (about 10 dpf) are exposed to high temperature for about 10 days (thermosensitivity varies among families); this sensitivity period also coincides with the gonad sensitivity period (Wessels and Hörstgen-Schwar 2007). It works by suppressing the aromatase gene (*Cyp19a1a*) activity, which subsequently results in low levels of estrogen; therefore the lack of estrogen results in the masculinisation of the genetic females. Suppression of the aromatase gene activity due to high temperature is also thought to occur through epigenetic modification, DNA methylation of the gene (Navarro-Martín *et al*. 2011).

It has also been reported that high temperatures, during the labile period of sex differentiation, induce a male biased sex-ratio by increasing the expression of *Dmrt1* and *Amh* genes followed by suppression of *Foxl2* gene expression and transcripts of the follicle

stimulating hormone receptor (FSHR) which suppresses the expression of the *Cyp19a1a* gene activity resulting in suppression of estrogen synthesis (Yamaguchi *et al*. 2007; Poonlaphdecha *et al*. 2013).

In thermosensitive species like tilapia, high temperature (> 34^{o}C) induces complete sex-reversal of XX female progeny into functional males in some families (Baroiller *et al*. 2009a) or a skewed sex-ratio in mixed sex groups (Kwon *et al.* 2002). In other species, high temperature induces a higher proportion of females in catfish, *Ictalurus punctatus* (Patiño *et al*. 1996) and sockeye salmon, *Oncorhynchus nerka* (Craig *et al*. 1996). Interestingly, an increasing proportion of females was observed when YY and XY Nile tilapia progeny (YY progeny were obtained from crossing YY males with YY females) were exposed to elevated temperatures (Abucay *et al*. 1999). Kwon *et al*. (2002) also found a significant feminisation effect (35.5 % female) in genetically YY Nile tilapia progeny and a very small proportion of females have also been found in a group of genetically XY fish when exposed to higher temperatures (36^{o}C).

The published feminisation protocols show very variable rates of feminisation and generally reduced survival in practice. Although in most cases elevated temperature treatments lead to masculinisation of genetic females, there are some evidences for feminisation of genetic males. The objective was to see if combining estrogen hormones (DES or EE2) and elevated temperature treatments could lead to efficient feminisation and survival rates similar to control groups despite the most common and opposite consequences of estrogen hormone and high temperature (estrogen induces feminisation and high temperature induces masculinisation in Nile tilapia). Experiments were set up with combined treatments of estrogen hormone and high temperature in mixed sex (obtained from crossing XX ♀ × XY♂), XY (from XX ♀ × YY♂) and XX (from XX ♀ × XX♂) progeny along with only

hormone or high temperature (positive controls) and a negative control at 28°C. Two different types of estrogen hormones (diethylstilbestrol, DES: 1000 mg/kg feed; 17α-ethynyloestradiol, EE2: 100 and 150 mg/kg feed) were used in order to determine the effectiveness of different feminising hormones in genetic male (XY) Nile tilapia. The genotypic sex of the fish was identified using tightly sex-linked SNP and microsatellite markers in LG1 and a SNP marker in LG20.

6.3 Materials and methods

6.3.1 Broodstock selection

The Nile tilapia (*Oreochromis niloticus*) stock, originating from Lake Manzala, Egypt, was maintained at the Institute of Aquaculture, University of Stirling, UK. A range of wild-type and red broodstock (19 male and 15 female wild-type; 23 male and 12 female red) were selected as potential broodstock (**Appendix Table C6.1**). Individual broodstock were tagged and a mucus sample was collected using filter paper (2 cm × 0.5 cm) followed by fixation in ethanol.

6.3.2 Genomic DNA extraction

Genomic DNA from filter paper (mucus) or fin clip (see below) was extracted using a salt-protein and isopropanol-DNA precipitation method modified from Aljanabi and Martinez (1997). In brief, a fin clip sample or mucus impregnated filter paper (the latter cut into small pieces) was digested overnight in SSTNE/SDS/proteinase K (10 mg/mL) at 55°C. Following RNaseA treatment (2 mg/mL) for one hour at 37°C, proteins were precipitated using 5 M NaCl (0.7 × vol.). The DNA was precipitated using equal volume of absolute isopropanol, dried and dissolved in 5 mM Tris. The quality and quantity of genomic DNA

was assessed using Nanodrop (ND-1000) spectrophotometer (Labtech International Ltd, UK) and 1 % agarose gel electrophoresis.

6.3.3 Genotypic sex identification

Individuals were genotyped with tightly sex-linked SNPs (*Oni23063*, *Oni28137*) and microsatellite (*UNH995*) markers in LG1 (Palaiokostas *et al*. 2013a) to identify the genotypic sex. Each individual was also genotyped for another sex-linked SNP marker (*Oni3161*) located in LG20 (Palaiokostas *et al*. 2015).

6.3.3.1 SNP marker genotyping

SNP assays were analysed using the fluorescence-based Kompetitive Allele Specific end point-PCR (KASP) genotyping system (LGC genomics Ltd UK, Semagn *et al*. 2014). The PCR reaction volume was 10 µL (c. 50 ng DNA) (primer sequences in **Appendix Table C4.4**) and the cyclic conditions were initial denaturation at 94^{o}C for 15 min, 10 cycles at 94^{o}C for 20 sec, 65^{o}C to 57^{o}C for 1 min (dropping 0.8^{o}C per cycle) and finally 34 cycles at 57^{o}C. Genotyping was based on the fluorescence signal detected by a Techne Quantica® machine (Barloworld Scientific Ltd. UK) and data was analysed using Quansoft software (version 1.1.21).

6.3.3.2 Microsatellite marker genotyping

The microsatellite marker (*UNH995*) was analysed using the fluorescent-labelled tailed primer method (Boutin-Ganache *et al*. 2001, primer sequences in **Appendix Table C4.5**). In short, 15 µL (c. 50 ng DNA) PCR reaction volumes were prepared and the thermal conditions were initial activation at 95^{o}C for 1 min and 35 cycles of denaturation at 95^{o}C for 15 sec, annealing at 62^{o}C for 15 sec and extension at 72^{o}C for 30 sec. PCR fragments were

then analysed on a CEQ^{TM} 8800 capillary sequencer (Beckman Coulter®, USA) according to the manufacturer's instructions and the alleles were annotated using the default parameters in CEQ software (version 9.0).

6.3.4 Crosses, and hormone and high temperature treatment

Following identification of the genotype of the broodstock, a range of crosses were produced XX♀ × XY♂; XX♀ × YY♂; XX♀ × XX♂ (**Table 6.1**) and each was treated with the different combinations of feminising hormones and high temperature.

Mature female and male broodstock were kept in glass aquaria and fed with commercially available trout feed (Trout Aquaculture Nutrition, UK; manufacturer Skretting, Preston, UK). Following ovulation, eggs were stripped manually and fertilised with fresh milt *in vitro*. Fertilised eggs were then incubated in down-welling incubators until the first feeding stage.

In total nine batches were produced from XX♀ × XY♂ crosses. The first two batches were subjected to a combined treatment of DES (1000 mg DES/kg feed) and high temperature (36^{o}C) with controls (28^{o}C/normal feed), with two replicates for each treatment. The other seven batches were divided into four different groups (the last two batches were divided into two groups only because of the lower number of fry) i.e. combined EE2 and high temperature (36^{o}C), EE2 (positive control), high temperature (36^{o}C, positive control) and 28^{o}C (normal feed; negative control). Combined EE2 and high temperature, and EE2 were replicated twice and no replication was used for single high temperature treatment because it was used as positive control for masculinisation. The first batch for EE2 experiment was treated at a dose of 100 mg EE2/kg feed and this was raised to 150 mg EE2/kg feed for the rest of the batches.

Table 6.1 Details of the cross type including the number of batches produced with the hormone treatment and dose.

Cross type	Progeny sex group	Number of batches produced	Hormone types	Dose
XX♀ × XY♂	Mixed (XX and XY)	2	DES	1000 mg/kg
XX♀ × XY♂		1	EE2	100 mg/kg
XX♀ × XY♂		6	EE2	150 mg/kg
XX♀ × YY♂	XY	1	DES and EE2	1000 mg/kg (DES) and 100 mg/kg (EE2)
XX♀ × YY♂		2	EE2	100 and 150 mg/kg
XX♀ × XX♂	XX	1	EE2	100 mg/kg

Three batches (XY spawns) were produced from XX♀ × YY♂ cross. The first batch was subjected to hormone treatment of DES (1000 mg/kg) and EE2 (100 mg/kg) with or without high temperature against a single treatment of only high temperature - 36^{o}C/normal feed and only control - 28^{o}C/normal feed (each group had two replicates). The remaining two batches were subjected to the different doses of EE2 hormone (100 and 150 mg EE2/kg respectively) with or without high temperature against a single treatment of high temperature and a control at 28^{o}C.

A single batch of XX fry was produced from an XX♀ × XX♂ cross and was treated with EE2 at a dose of 100 mg/kg feed. This batch was also split into four different groups: EE2 with or without high temperature against a single treatment of high temperature and control.

Batches were produced at different times, as the broodstock did not spawn at the same time; the capacity of the temperature-controlled troughs was also limited (one each at 36 and 28^{o}C, capacity 10 plastic aquaria per trough). Hormone-treated feed was prepared by dissolving the required amount of hormone (DES/EE2) into 100 % ethanol and then mixing this with the required amount of finely ground fry feed. The hormone feed was air dried to evaporate the ethanol and kept in dry, cool condition until use (Guerrero III 1975).

Once the fry (swim up stage, approx.10 dpf) were ready for the treatment, an equal number of fry (ranged from 31 to 70) were counted for each treatment group from each batch and each group was transferred into a plastic aquarium with 5 L of water where the water temperature was same as the incubation temperature (28^{o}C). The aquaria were then placed into a temperature-controlled trough, either at 36 or 28^{o}C, with a thermostatically controlled heater and pumps to circulate water around the tanks so all tanks in each trough were maintained at identical temperatures. The water temperature for control and high temperature treatment were 27.96^{o}C ± 0.03 and 36.11^{o}C ± 0.12 (mean ± SD) respectively throughout the experiments. For hormone treatments at 28^{o}C, fry were fed with hormone-treated feed five times a day *ad libitum* for 20 days. For the combined treatment of hormone and high temperature, aquaria were placed into the 36^{o}C trough so that the water temperature inside the trough increased gradually from 28^{o}C to 36^{o}C to acclimatise the fry. Following 10 days of the hormone and high temperature treatment, the aquarium was moved to the 28^{o}C trough where the temperature decreased gradually to 28^{o}C and hormone treatment was continued for another 10 days. For high temperature treatment without hormone, the water temperature was maintained at 36^{o}C for 10 days and the fry were fed with normal feed (followed by transfer to 28^{o}C). The fry in the negative control group from each batch were fed with normal feed and maintained at 28^{o}C. Leftover feed was siphoned out and the water temperature was checked twice a day. Once the treatment finished, the number of fry in each group was counted to calculate the survival rate and they were transferred to 20 L tanks in a recirculating water system until they were large enough for phenotypic sex identification.

6.3.4.1 Phenotypic and genotypic sex identification

After 90 days of rearing, phenotypic sex was determined using the acetocarmine gonad squash method (Guerrero III and Shelton 1974) and a fin clip was preserved in 100 % ethanol for genotypic sex identification. For gonad squash, fish were killed by an overdose of benzocaine (Sigma Aldrich, UK) and destruction of the central nervous system (Schedule 1 method) and the whole gonad was microscopically examined. The number of sexed fish per cross ranged from 14 - 67 for XX♀ × XY♂ crosses, 15 - 60 for XX♀ × YY♂ crosses and 35 - 40 for the XX♀ × XX♂ cross. One replicate from each hormone treatment and combined hormone/high temperature treatment from the batch (where possible) were retained to identify XY neo-females as future broodstock.

Genomic DNA was extracted from the samples from each cross and were analysed for LG1 (2 SNPs and 1 microsatellite) and LG20 markers using the protocol described in section 6.3.2 and 6.3.3. Two SNPs in LG1 (*Oni23063, Oni28137*) gave 100 % similar results in the broodstock and first two DES treated batches (XX♀ × XY♂ cross). The microsatellite marker was also analysed for the broodstock, and first two DES and two EE2 treated batches (XX♀ × XY♂ cross). The genotypes were in agreement with the SNPs (with very few exceptions), therefore only one SNP (*Oni23063*) was analysed from LG1 for the rest of the batches. If both parents were homozygous for the same allele at the LG20 SNP marker (*Oni3161*) for any of the batches, progeny were not tested for that marker.

6.3.4.2 XY neo-female identification as future broodstock

Once the fishes were big enough to tag (about 6-8 months old), phenotypic sex was identified, they were PIT-tagged and a mucus sample was collected using filter paper for genotypic sex identification. Following DNA extraction, individuals were analysed for LG1

(*Oni23063*) and LG20 (*Oni3161*) SNPs following the protocol described in section 6.3.2 and 6.3.3, to identify the genotypic sex.

6.3.5 Data analysis

Survival rates were expressed as a percentage. Survival rates were assessed within the replicates using initial and final number after the treatment for homogeneity test in 2 × 2 contingency table using the G-test of independence (maximum likelihood statistical significance test) followed by comparing the significance level between the pooled replicates for each treatment against the values of their corresponding control or pooled controls. Phenotypic sex-ratio and the genotypic segregation data for the control group were initially assessed by Chi-square goodness of fit test (any deviation from expected 1:1 ratio) before conducting further statistical tests. Sex-ratios of treated groups were compared with the control groups using 2 × 2 contingency G-test of independence. Phenotypic and genotypic sex reversed neo-female/neo-male were expressed as percentage over the total number of fish analysed. A probability value of $p < 0.05$ was considered as significant.

6.4 Results

6.4.1 Analysis of broodstock with sex-linked markers

Males were heterozygotes (A/G for *Oni23063* and G/T for *Oni28137*) for LG1 SNP markers whereas females were homozygotes (G/G and T/T respectively) for those SNP markers, with the exception of 12 males (**Table 6.2**; details in **Appendix Table C6.1**). Nine of the males (6 from wild-type and 3 from red strain) were homozygotes for the G allele for *Oni23063* marker and the T allele for *Oni28137* marker, and those were classified as putative XX neo-males. Three males (red strain) were homozygotes for the A allele for

Oni23063 marker and the G allele for *Oni28137* marker (**Table 6.2**), and these were classified as putative YY males. Based on the LG1 microsatellite marker, the fish with allele 232 or 236 were always classified as XY or YY males based on the LG1 SNPs; no phenotypic female was identified with either of these two alleles. Homozygotes for allele 184 were always associated with females (**Table 6.2**). The LG20 marker genotypes did not show any association with the phenotypic sex of the wild-type broodstock analysed, whereas significant linkage was found between the genotype and phenotypic sex in the red-type broodstock (**Table 6.2**).

Table 6.2 Allelic combinations of the SNPs (*Oni23063, Oni28137*) and microsatellite (*UNH995*) markers in LG1 and a SNP marker in LG20 (*Oni3161*) in the male and female broodstock for wild-type and red strain.

Wild-type strain												
LG1									**LG20**			
SNPs						**Microsatellite**			**SNP**			
Oni23063			***Oni28137***			***UNH995***			***Oni3161***			
Genotype	**Male**	**Female**	**Genotype**	**Male**	**Female**	**Genotype**	**Male**	**Female**	**Genotype**	**Male**	**Female**	***P*-value**
A/A	0	0	G/G	0	0	184/184	1	1	C/C	5	2	0.33
A/G	13	0	G/T	13	0	184/188	4	2	C/T	4	7	
G/G	6	15	T/T	6	15	184/190	0	1	T/T	10	6	
						188/188	2	0				
						188/190	2	2				
						188/192	3	0				
						190/192	0	3				
						184/224	0	1				
						188/224	4	1				
						192/224	2	2				
						224/224	1	1				

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

Table 6.2 (Cont'd) Allelic combinations of the SNPs (*Oni23063*, *Oni28137*) and microsatellite (*UNH995*) markers in LG1 and a SNP marker in LG20 (*Oni3161*) to the male and female broodstock for wild-type and red strain.

Red strain												
LG1									**LG20**			
SNPs						**Microsatellite**			**SNP**			
Oni23063			***Oni28137***			***UNH995***			***Oni3161***			
Genotype	**Male**	**Female**	**Genotype**	**Male**	**Female**	**Genotype**	**Male**	**Female**	**Genotype**	**Male**	**Female**	***P*-value**
A/A	3	0	G/G	3	0	184/184	3	7	C/C	0	3	0.03
A/G	17	0	G/T	17	0	184/188	1	5	C/T	14	4	
G/G	3	12	T/T	3	12	188/188	1	0	T/T	9	5	
						184/236	6	0				
						188/236	6	0				
						232/232	2	0				
						236/236	1	0				
						232/252	2	0				
						236/252	1	0				

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

6.4.2 Mixed sex group experiments (XX♀ × XY♂)

The survival rates in normal cross ranged from 24 % - 55 % in the first two batches of combined DES and high temperature treatments compared to 93 - 100 % for the control groups (**Table 6.3; Figure 6.1**). The overall survival rates were significantly lower in treatment groups than in their respective controls (*p*-value 9.38e-11 and 7.78e-03 for the two batches respectively), however there was no significant difference between the replicates in treatment and control groups (**Table 6.3**).

A significant level of phenotypic sex-reversal (male to female) was observed in combined DES and high temperature treated batches compared to the control (*p*-value 4.45e-05 and 5.50e-12 for 2 batches respectively, **Table 6.3**). There was no deviation from the expected phenotypic sex-ratio (1:1) in the first control batch (*p*-value 0.99) whereas a significant deviation was found in the other control batch (*p*-value 9.0e-04).

All the samples from first two batches (treated with DES) were tested for LG1 and LG20 markers. The combined DES and high temperature treatment group had significantly higher mortality than the untreated control group; the genotypic data from the survivors suggested that there was no deviation from the expected ratio. Therefore there did not appear to be differential mortality among genotypes in this treatment. LG1 markers (both SNPs, *Oni23063* and *Oni28137*; and microsatellite, *UNH995*) suggest that 42 - 46 % males (genotypically XY) changed their phenotypic sex to female in combined DES and high temperature treatments (**Table 6.4; Figure 6.2**). There was no deviation of genotype segregation from the expected ratio in the control groups; males were heterozygotes and females were homozygotes for the SNPs in LG1 (*Oni23063*, *Oni28137*) (**Table 6.4**). Deviation from the expected genotype was observed in two cases for *UNH995* marker (**Table 6.4**). In batch 2 where there was a significant deviation from the expected

phenotypic sex-ratio observed in control group, genotyping results showed that a significant proportion of the phenotypic females (*p*-value 2.52e-06) had the expected male-linked genotype (XY) for the sex-linked markers in LG1 (*Oni23063* and *UNH995*).

Table 6.3 Survival rate of Nile tilapia from XX♀ × XY♂ cross after 20 days of treatment with DES (1000 mg/kg) and high temperature, with 28°C (no hormone) as negative control, and the phenotypic sex-ratio after three months of rearing at 28°C.

Batch (Hormone dose)	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
1 (1000 mg/kg)	28°C	1	65	65	100	0.99		58	29	29	0	50	
		2	65	65	100								
	Mean ± SD				**100 ± 0.0**								
	DES + 36°C	1	65	14	21.54	0.628	9.38e-11						
		2	65	17	26.15			14	0	14	0	0	4.45e-05
	Mean ± SD				**23.85 ± 2.31**								
2 (1000 mg/kg)	28°C	1	64	60	93.75	0.95		52	38	14	0	73.08	
		2	64	59	92.19								
	Mean ± SD				**92.97 ± 0.78**								
	DES + 36°C	1	64	39	60.94	0.50	7.78e-03	27	0	26	1	0	5.50e-12
		2	64	32	50								
	Mean ± SD				**55.47 ± 5.47**								

Table 6.4 Allelic inheritance of SNPs (*Oni23063*, *Oni28137*) and microsatellite (*UNH995*) markers in LG1 and a SNP (*Oni3161*) in LG20 to the progeny from XX♀ × XY♂ cross treated with DES (1000 mg/kg) and high temperature (36°C) including control (28°C), and the percentage of neo-female (genotypically XY) after three months of rearing.

		LG1												**LG20**			
		SNPs								**Microsatellite**				**SNP**			
Batch	**Treatment**	***Oni23063***				***Oni28137***				***UNH995***				***Oni3161***			
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Genotype**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	***P*-value**
1	28°C	G/G	0	29	0.0	T/T	0	29	0.0	184/188	10	1	4.35	T/T	6	5	0.24
		A/G	29	0		G/T	29	0		188/188	18	0		C/T	13	19	
		A/A	0	0		G/G	0	0		184/192	1	14		C/C	10	5	
										188/192	0	14					
	DES + 36°C	G/G	0	7	46.15	T/T	0	7	46.15	184/188	0	3	46.15	T/T	0	1	0.99
		A/G	0	6		G/T	0	6		188/188	0	4		C/T	0	8	
		A/A	0	0		G/G	0	0		184/192	0	3		C/C	0	4	
										188/192	0	3					
2	28°C	G/G	11	14	0.0					184/232	27	0	0.0	T/T	24	8	0.75
		A/G	27	0						184/252	11	13		C/T	14	6	
	DES + 36°C	G/G	0	15	42.31					184/232	0	12	42.85	T/T	0	17	0.99
		A/G	0	11						184/252	0	14		C/T	0	9	

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

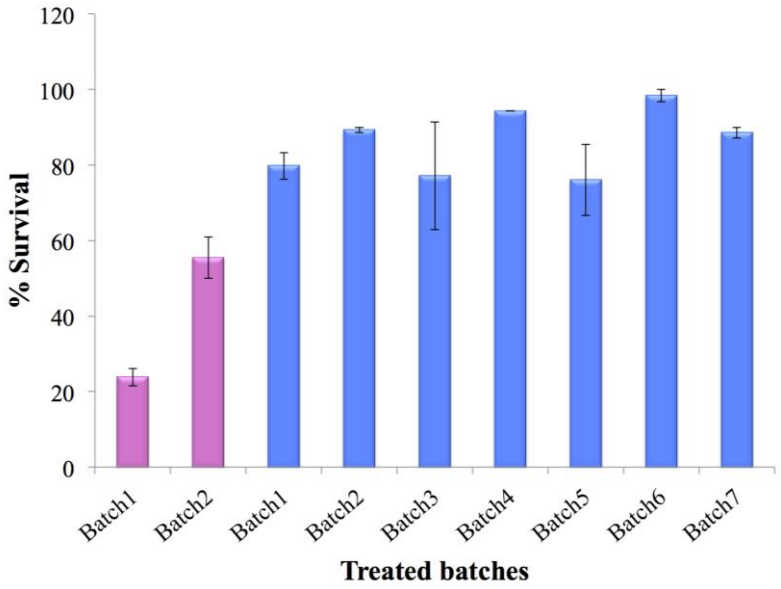


Figure 6.1 Survival rate (mean ± SD) of the fry immediately after 20 days of combined treatment with hormone and high temperature from the XX♀ × XY♂ crosses. The X-axis represents the batches produced and the Y-axis represents the percent survival after the treatment. DES + 36^{o}C - purple colour bar, EE2 + 36^{o}C - blue colour bar.

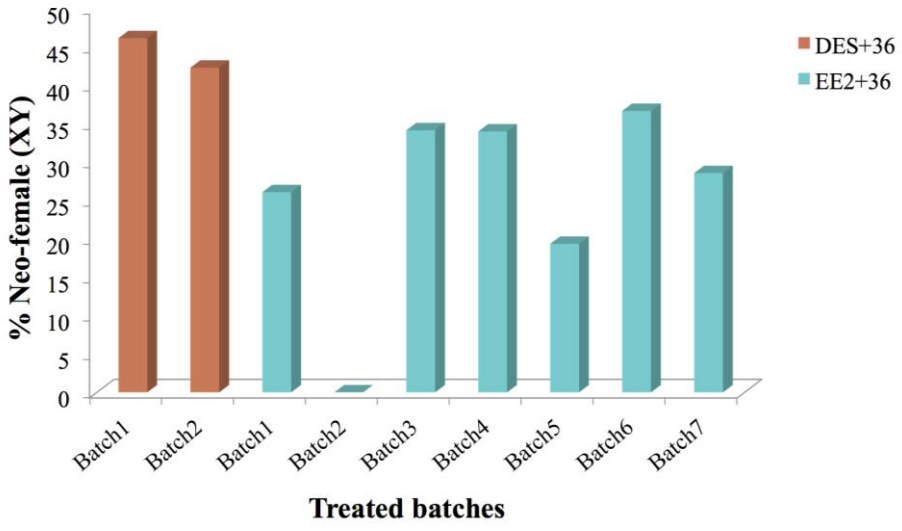


Figure 6.2 Percentage of XY neo-females (phenotypically female but genotypically male) in combined treatment of hormone and high temperature from the XX♀ × XY♂ cross after gonad squash based on the LG1 marker (*Oni23063*). The X-axis represents the batches produced and the Y-axis represents the percentage of neo-females.

In all EE2 treated batches except batch 2 (each batch had four groups where possible; combined EE2 and high temperature, EE2, high temperature and control at 28^{o}C), survival rates ranged from 62 - 100 % and 79 - 100 % in treated and control groups respectively (**Figure 6.1; Table 6.5**). There were no significant differences in survival rates between each treatment (combined EE2 and high temperature, EE2, high temperature) and control groups or between the replicates (**Table 6.5**). Only one batch (batch 2) showed significant differences in survival rates between the treatment and the control. These differences did not appear to be related to the hormone or temperature treatment because the control also had a lower survival rate (52.86 % mortality) compared to all the other treatments and the major mortalities occurred at the beginning of the experiment (**Table 6.5**).

A significant proportion of sex-reversal (based on gonad squash) was found in combined EE2 and high temperature induced batches (except batch 2) irrespective of the doses used compared to the control (no deviation from the expected 1:1 ratio at 28^{o}C, **Table 6.5**). EE2 hormone alone showed no significant phenotypic sex inversion in the batches treated except in batch 5, which showed significant feminisation rate against control (*p*-value 2.00e-04, **Table 6.5**). High temperature induced masculinisation was observed in only one of the treated batches (Batch 4), where a significant proportion of the progeny had converted their sex from female to male compared to the control (*p*-value 1.89e-02, **Table 6.5**).

Based on the sex-linked marker analysis, 19 - 37 % of the fish were found to be genetically XY neo-females in combined EE2 and high temperature induced batches (except batch 2, **Figure 6.2**), and the observed genotypic ratio was same as the expected ratio in the control groups (28^{o}C) for all the batches (**Table 6.6**). Homozygotes for the LG1 SNP (*Oni23063*) were always associated with female and heterozygotes with males from all the batches (out of 277 genotyped samples) except one male in control (batch 1). That male also showed the

female genotype for the *UNH995* marker. In the EE2 induced group (batch 5) where a significant sex-reversal was observed based on the phenotypic sex (after gonad squash), a proportion of the genetic males (39 %) were found to have been converted to phenotypic females based on the sex-linked marker analysis (**Table 6.6**). High temperature showed a significant masculinisation rate in only one of the treated batches based on the phenotypic sex (batch 4), whereas marker analysis showed 9 - 17 % of genetic females (XX) converted their phenotypic sex to male (neo-male) in three of the batches (**Table 6.6; Figure 6.3**). From the phenotypic sex (after gonad squash) and genotyping sex data (using sex-linked markers), none of the treatments in batch 2 showed any phenotypic or genotypic sex deviation (**Table 6.5, 6.6**).

One replicate from each group was kept alive to identify XY neo-females as potential future broodstock. Each fish was analysed for sex-linked marker to identify XY neo-females. The 20 - 36 % and 15 - 68 % neo-females were identified from combined DES and EE2 with high temperature treated batches respectively, and 15 % from one of the EE2 treated batches (batch 5, which had significant difference in phenotypic sex-ratio, **Table 6.7; Figure 6.4**). No XY neo-female was identified from batch 1 with combined EE2 and high temperature treatment, which had high mortalities (**Table 6.7**).

The results from the SNP marker (*Oni3161*) in LG20 suggest that there was no evidence of LG20 influence on the phenotypic sex in any of the batches, even for the batch where the phenotypic sex in the control group was significantly deviated from the expected ratio (batch 2 from XX♀ × XY♂ cross) (**Table 6.4**, **6.6, 6.7**).

Table 6.5 Survival rate of Nile tilapia from XX♀ × XY♂ cross after 20 days of treatment for EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the phenotypic sex-ratio after three months of rearing at 28°C.

Batch (Hormone dose)	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
1 (100 mg/kg)	28°C	1	42	33	78.57			20	9	11	0	45	
	36°C	1	42	26	61.9		0.48	24	11	13	0	45.83	0.96
	EE2	1	42	34	80.95	0.86	0.84	23	9	14	0	39.13	0.70
		2	42	36	85.71								
	Mean ± SD				**83.33 ± 2.38**								
	EE2 + 36°C	1	42	32	76.19	0.78	0.96	28	2	21	5	7.14	5.30e-03
		2	42	35	83.33								
	Mean ± SD				**79.76 ± 3.57**								
2 (150 mg/kg)	28°C	1	70	33	47.14			23	8	15	0	34.78	
	36°C	1	70	63	90		0.02	56	22	34	0	39.29	0.71
	EE2	1	70	58	82.85	1.88e-05	0.97	14	4	9	1	28.57	0.81
		2	70	15	21.43								
	Mean ± SD				**52.14 ± 30.71**								
	EE2 + 36°C	1	70	63	90	0.95	7.84e-03	59	15	42	2	25.42	0.45
		2	70	62	88.57								
	Mean ± SD				**89.29 ± 0.72**								
3 (150 mg/kg)	28°C	1	70	61	87.14			57	31	26	0	54.39	
	36°C	1	70	68	97.14		0.66	67	36	31	0	53.73	0.94
	EE2	1	70	55	78.57	0.68	0.82	47	21	26	0	44.68	0.32
		2	70	61	87.14								
	Mean ± SD				**82.86 ± 4.29**								
	EE2 + 36°C	1	70	44	62.86	0.15	0.57	42	5	36	1	11.9	8.20e-06
		2	70	64	91.43								
	Mean ± SD				**77.15 ± 14.29**								

Table 6.5 (Cont'd) Survival rate of Nile tilapia from XX♀ × XY♂ cross after 20 days of treatment for EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the phenotypic sex-ratio after three months of rearing at 28°C.

Batch (Hormone dose)	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
4 (150 mg/kg)	28°C	1	70	67	95.71			44	23	21	0	52.27	
	36°C	1	70	67	95.71			59	44	15	0	74.57	1.89e-02
	EE2	1	70	64	91.43	0.90	0.77	41	18	23	0	43.9	0.44
		2	70	62	88.57								
	Mean ± SD				**90 ± 1.43**								
	EE2 + 36°C	1	70	66	94.29	0.99		50	9	41	0	18	4.00e-04
		2	70	66	94.29								
	Mean ± SD				**94.29 ± 0.0**								
5 (150 mg/kg)	28°C	1	48	48	100			48	27	21	0	56.25	
	36°C	1	48	48	100			28	12	16	0	42.86	0.26
	EE2	1	48	33	68.75	0.31	0.42	31	5	26	0	16.13	2.00e-04
		2	48	45	93.75								
	Mean ± SD				**81.25 ± 12.5**								
	EE2 + 36°C	1	48	32	66.67	0.43	0.29	31	8	23	0	25.81	6.90e-03
		2	48	41	85.42								
	Mean ± SD				**76.05 ± 9.38**								
6 (150 mg/kg)	28°C	1	31	28	90.32			27	15	12	0	55.56	
	EE2 + 36°C	1	31	31	100	0.93	0.79	30	1	29	0	3.33	3.01e-06
		2	31	30	96.77								
	Mean ± SD				**98.39 ± 1.62**								
7 (150 mg/kg)	28°C	1	70	70	100			59	27	32	0	45.76	
	EE2 + 36°C	1	70	61	87.14	0.90	0.56	56	11	45	0	19.64	2.60e-03
		2	70	63	90								
	Mean ± SD				**88.57 ± 1.43**								

Table 6.6 Allelic inheritance of SNP (*Oni23063*) and microsatellite (*UNH995*) markers in LG1 and a SNP (*Oni3161*) in LG20 to the progeny from XX♀ × XY♂ cross treated with EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the percentage of neo-female (genotypically XY) after three months of rearing.

		LG1								**LG20**			
		SNP				**Microsatellite**				**SNP**			
Batch (Hormone dose)	**Treatment**	***Oni23063***				***UNH995***				***Oni3161***			
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	***P*-value**
1 (100 mg/kg)	28°C	G/G	1	11	0.0	184/184	0	6	0.0	T/T	4	5	0.99
		A/G	8	0		184/188	1	5		C/T	5	6	
						184/236	4	0					
						188/236	4	0					
	36°C	G/G	3	13	0.0	184/184	2	8	0.0	T/T	9	7	0.21
		A/G	8	0		184/188	2	5		C/T	2	6	
						184/236	5	0					
						188/236	2	0					
	EE2	G/G	0	14	0.0	184/184	0	7	4.35	T/T	5	9	0.99
		A/G	9	0		184/188	0	6		C/T	4	5	
						184/236	6	0					
						188/236	3	1					
	EE2 + 36°C	G/G	0	15	26.09	184/184	0	7	26.09	T/T	0	11	0.48
		A/G	2	6		184/188	0	8		C/T	2	10	
						184/236	2	2					
						188/236	0	4		T/T	4	5	0.99

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

Table 6.6 (Cont'd) Allelic inheritance of SNP (*Oni23063*) and microsatellite (*UNH995*) markers in LG1 and a SNP (*Oni3161*) in LG20 to the progeny from XX♀ × XY♂ cross treated with EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the percentage of neo-female (genotypically XY) after three months of rearing.

		LG1								**LG20**		
		SNP				**Microsatellite**				**SNP**		
Batch (Hormone dose)	**Treatment**	***Oni23063***				***UNH995***				***Oni3161***		
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**
2 (150 mg/kg)	28°C	G/G	0	15	0.0	184/184	0	7	0.0	T/T		
		A/G	8	0		184/188	0	8		T/T		
						184/236	4	0				
						188/236	4	0				
	36°C	G/G	0	34	0.0	184/184	0	17	1.79			
		A/G	22	0		184/188	0	16				
						184/236	11	1				
						188/236	11	0				
	EE2	G/G	0	9	0.0	184/184	0	5	0.0			
		A/G	4	0		184/188	0	4				
						184/236	1	0				
						188/236	3	0				
	EE2 + 36°C	G/G	0	42	0.0	184/184	1	19	3.51			
		A/G	15	0		184/188	1	21				
						184/236	5	1				
						188/236	8	1				

Table 6.6 (Cont'd) Allelic inheritance of SNP (*Oni23063*) and microsatellite (*UNH995*) markers in LG1 and a SNP (*Oni3161*) in LG20 to the progeny from XX♀ × XY♂ cross treated with EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the percentage of neo-female (genotypically XY) after three months of rearing.

		LG1				**LG20**			
		SNP							
Batch (Hormone dose)	**Treatment**	***Oni23063***				***Oni3161***			
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	***P*-value**
3 (150 mg/kg)	28°C	G/G	0	26	0.0	T/T	7	7	0.94
		A/G	30	0		C/T	7	6	
						C/C	17	13	
	36°C	G/G	6	31	0.0	T/T	5	6	0.10
		A/G	30	0		C/T	8	13	
						C/C	23	12	
	EE2	G/G	0	25	2.13	T/T	8	3	0.08
		A/G	21	1		C/T	7	9	
						C/C	6	14	
	EE2 + 36°C	G/G	0	22	34.15	T/T	2	9	0.84
		A/G	5	14		C/T	1	10	
						C/C	2	17	
4 (150 mg/kg)	28°C	G/G	0	21	0.0	T/T	0	0	0.77
		A/G	23	0		C/T	13	13	
						C/C	10	8	
	36°C	G/G	10	15	0.0	T/T	0	0	0.23
		A/G	34	0		C/T	20	10	
						C/C	24	5	
	EE2	G/G	0	23	0.0	T/T	0	0	0.75
		A/G	18	0		C/T	7	11	
						C/C	11	12	
	EE2 + 36°C	G/G	0	24	34.0	T/T	0	0	0.99
		A/G	9	17		C/T	5	21	
						C/C	4	20	
5 (150 mg/kg)	28°C	G/G	0	21	0.0	T/T	14	12	0.78
		A/G	27	0		C/T	13	9	
	36°C	G/G	0	15	0.0	T/T	6	11	0.44
		A/G	13	0		C/T	6	5	
	EE2	G/G	0	14	38.71	T/T	3	13	0.99
		A/G	5	12		C/T	2	13	
	EE2 + 36°C	G/G	0	17	19.35	T/T	5	9	0.41
		A/G	8	6		C/T	3	14	

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

Table 6.6 (Cont'd) Allelic inheritance of SNP (*Oni23063*) and microsatellite (*UNH995*) markers in LG1 and a SNP (*Oni3161*) in LG20 to the progeny from XX♀ × XY♂ cross treated with EE2 (100 and 150 mg/kg) with or without high temperature including 28°C and 36°C as negative and positive controls respectively, and the percentage of neo-female (genotypically XY) after three months of rearing.

		LG1				**LG20**			
		SNP							
Batch (Hormone dose)	**Treatment**	***Oni23063***				***Oni3161***			
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	***P*-value**
6 (150 mg/kg)	28°C	G/G	0	12	0.0	C/C			
		A/G	15	0		C/C			
	EE2 + 36°C	G/G	0	18	36.67				
		A/G	1	11					
7 (150 mg/kg)	28°C	G/G	0	32	0.0	T/T	15	13	0.30
		A/G	27	0		C/T	12	19	
	EE2 + 36°C	G/G	0	29	28.57	T/T	6	21	0.74
		A/G	11	16		C/T	5	24	

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

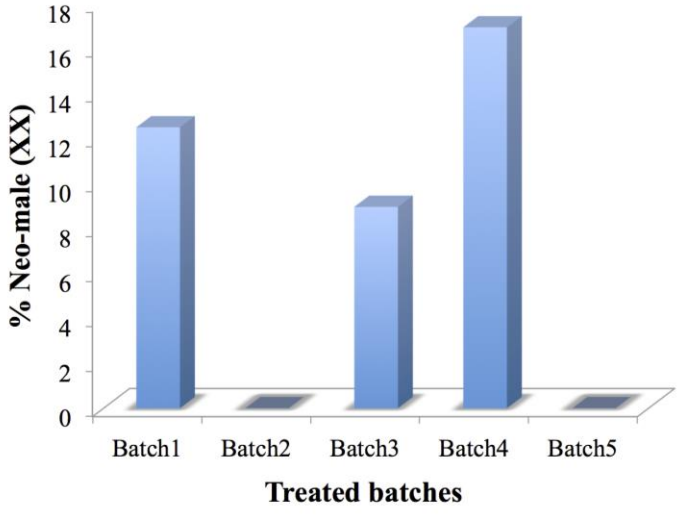


Figure 6.3 Percentage of neo-male (XX, phenotypically male but genotypically female) in the batches treated with high temperature from the XX♀ × XY♂ cross following gonad squash. Male which has XX genotype (female genotype) based on LG1 marker (*Oni23063*) is considered as neo-male. X-axis represents the number of batches and Y-axis represents the percent neo-male.

Table 6.7 Allelic inheritance of SNP markers in LG1 (*Oni23063*) and LG20 (*Oni3161*) to the progeny from XX♀ × XY♂ cross treated for DES and EE2 with or without high temperature, and the percentage of neo-female (genotypically XY) after 6-8 months of rearing for future broodstock.

		LG1				**LG20**			
		SNP							
Batches	**Treatment**	***Oni23063***				***Oni3161***			
		Genotype	**Male**	**Female**	**Neo-female (%)**	**Genotype**	**Male**	**Female**	***P*-value**
1	DES + 36	G/G	1	2	20.0	T/T	1	2	0.74
		A/G	5	2		C/T	3	1	
						C/C	2	1	
2	DES + 36	G/G	1	12	36.36	T/T	1	9	0.22
		A/G	8	12		C/T	8	15	
1	EE2 + 36	G/G	0	5	0.00	T/T	6	3	0.52
		A/G	7	0		C/T	1	2	
3	EE2	G/G	0	24	0.00	T/T	5	6	0.23
		A/G	10	0		C/T	1	9	
						C/C	4	9	
3	EE2 + 36	G/G	0	25	24.07	T/T	5	9	0.86
		A/G	16	13		C/T	3	10	
						C/C	8	19	
4	EE2 + 36	G/G	0	29	22.64	C/T	9	15	0.12
		A/G	12	12		C/C	5	24	
5	EE2	G/G	0	17	15.15	T/T	4	12	0.46
		A/G	11	5		C/T	7	10	
5	EE2 + 36	G/G	0	19	14.71	T/T	4	6	0.43
		A/G	10	5		C/T	6	18	
6	EE2 + 36	G/G	0	8	67.86				
		A/G	1	19					
7	EE2 + 36	G/G	0	23	40.00	T/T	2	20	0.27
		A/G	7	20		C/T	7	21	

P-value represents the association between genotype for LG20 marker and the phenotypic sex (Fisher's exact test)

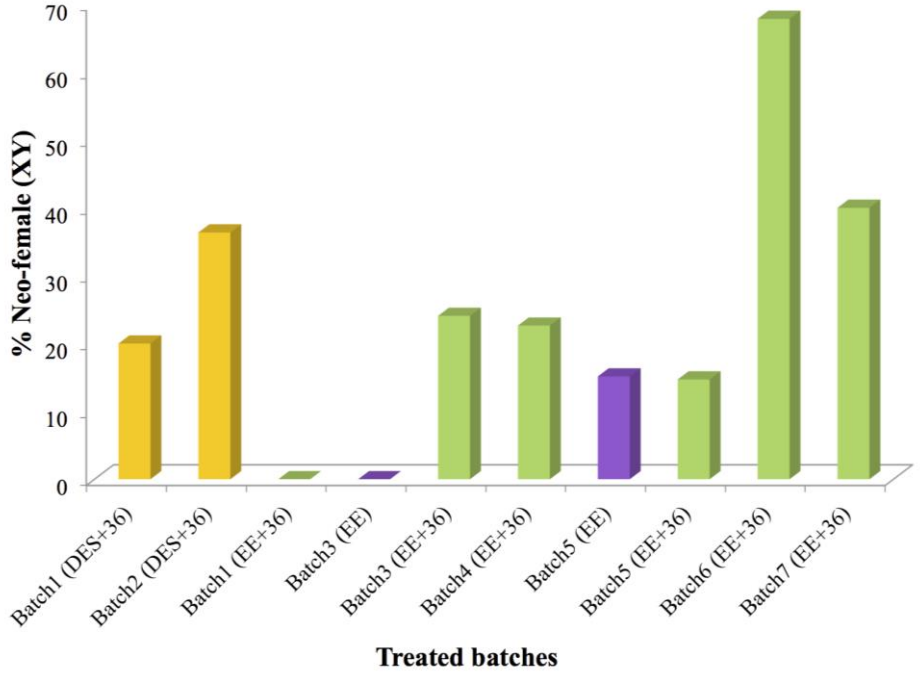


Figure 6.4 Percentage of neo-female (phenotypically female but genotypically male) in the batches with combined treatment of hormone and high temperature from the XX♀ × XY♂ cross after 6-8 months of rearing as future broodstock. Female which has XY genotype (male genotype) based on LG1 marker (*Oni23063*) is considered as neo-female. X-axis represents the number of batches produced and Y-axis represents the percent neo-female.

6.4.3 XY sex group experiment (XX♀ × YY♂)

The first batch from XX♀ × YY♂ cross was subjected to both DES (1000 mg/kg) and EE2 (100 mg/kg) hormone with or without high temperature. The survival rate ranged from 33 - 90 % (**Figure 6.5**). There was no significant difference in survival rates between the replicates or between the treatment and control groups in EE2 treated batches. The results were same for the replicates in DES treated groups but a statistically significant increase in mortality was observed between the DES treatment and control (**Table 6.8**). All the progeny were males in the control group (28^{o}C) as expected from an XX♀ × YY♂ cross. In the high

temperature group (36^{o}C), 100 % of the progeny were male. A very high proportion (77 - 94 %, statistically significant) of genetic males changed their phenotypic sex to female, slightly higher in the DES group than for EE2 (**Table 6.8**). The combined treatment of hormone and high temperature showed higher sex-reversal rate (89 - 94 %) compared to the hormone alone (77 - 84 %, **Table 6.8**).

There was no significant difference in survival rate between the replicates or between the treatment and control, and the survival rate ranged from 74 - 96 % for the other two batches from XX♀ × YY♂ crosses (**Table 6.9**). All the fish in control and high temperature treatments were male. A statistically significant number of males (8 - 18 %) changed their phenotypic sex to female in combined EE2 and high temperature treatment compared to the 2 % observed in only EE2 treatment (**Table 6.9**). All the progeny from the combined EE2 and high temperature treatment, and 15 progeny from other groups were tested with sex-linked markers to confirm their genotypic sex and all the genotypes were XY as expected. High temperature did not induce any feminisation of the XY group (**Table 6.8, 6.9**).

Table 6.8 Survival rate of Nile tilapia from XX♀ × YY♂ cross after 20 days of treatment for DES (1000 mg/kg) and EE2 (100 mg/kg) with or without high temperature including 28°C and 36°C, and the sex-ratio after three months of rearing at 28°C.

Batch	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
1	28°C	1	50	46	92	0.88		46	46	0	0	100	
		2	50	44	88			44	44	0	0	100	
	Mean ± SD				**90 ± 2.0**							**100 ± 0**	
	36°C	1	50	44	88	0.93	0.96	44	44	0	0	100	
		2	50	45	90			45	45	0	0	100	
	Mean ± SD				**89 ± 1.0**							**100 ± 0**	**0.99**
	DES	1	50	20	40	0.80	9.63e-04	20	3	17	0	15.00	
		2	50	22	44			22	4	18	0	18.18	
	Mean ± SD				**42 ± 2.0**							**16.59 ± 1.59**	**3.52e-14**
	DES + 36°C	1	50	18	36	0.65	3.13e-05	18	1	17	0	5.56	
		2	50	15	30			15	1	13	1	6.67	
	Mean ± SD				**33 ± 3.0**							**6.12 ± 0.56**	**7.11e-15**
	EE2	1	50	42	84	0.94	0.78	42	10	32	0	23.81	
		2	50	43	86			43	10	32	1	23.26	
	Mean ± SD				**85 ± 1.0**							**23.54 ± 0.28**	**2.2e-16**
	EE2 + 36°C	1	50	40	80	0.75	0.74	40	5	34	1	12.5	
		2	50	44	88			44	4	40	0	9.09	
	Mean ± SD				**84 ± 4.0**							**10.8 ± 1.71**	**2.2e-16**

Table 6.9 Survival rate of Nile tilapia from XX♀ × YY♂ cross after 20 days of treatment for EE2 with or without high temperature (100 and 150 mg/kg) including 28°C and 36°C, and the sex-ratio after three months of rearing at 28°C.

Batch (Hormone dose)	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
2 (100 mg/kg)	28°C	1	70	65	92.85			56	56	0	0	100	
	36°C	1	70	65	92.86			60	60	0	0	100	0.99
	EE2	1	70	68	97.14	0.90	0.89	49	47	1	1	95.92	0.21
		2	70	66	94.29								
	Mean ± SD				**95.72 ± 1.43**								
	EE2 + 36°C	1	70	56	80	0.64	0.68	50	38	9	3	76	9.89e-05
		2	70	63	90								
	Mean ± SD				**85 ± 5**								
3 (150 mg/kg)	28°C	1	70	66	94.29			50	50	0	0	100	
	36°C	1	70	67	95.71		0.96	45	45	0	0	100	0.99
	EE2	1	70	52	74.29	0.94	0.25	29	29	0	0	100	0.99
		2	70	51	72.86								
	Mean ± SD				**73.58 ± 0.72**								
	EE2 + 36°C	1	70	53	75.71	0.49	0.54	49	45	4		91.84	1.60e-02
		2	70	63	90								
	Mean ± SD				**82.86 ± 7.15**								

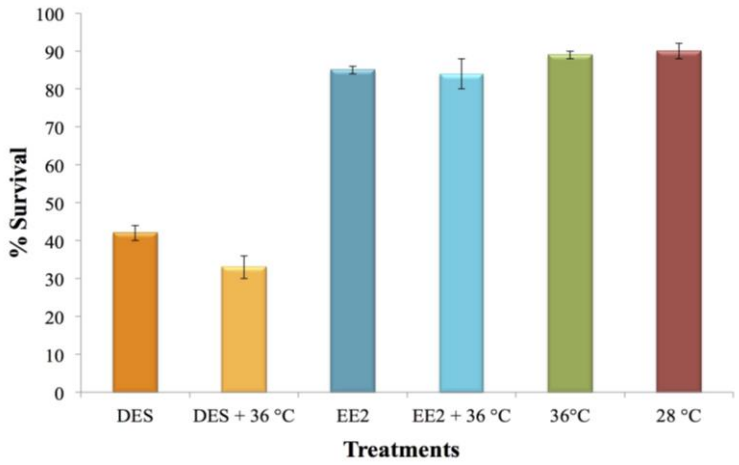


Figure 6.5 Survival rate (Mean ± SD) of the fry immediately after 20 days of treatment from the batch 1 (XX♀ × YY♂ cross). X-axis represents the treatments including control and Y-axis represents the percent survival.

6.4.4 XX sex group experiment (XX♀ × XX♂)

Survival rate was > 83 % in treated and control groups, and there were no significant differences in the survival rate between the replicates or between treatment and control (**Table 6.10**). Approximately 9 % of the progeny were male in the control group where 100 % female progeny were expected, whereas 2.7 % and 0 % males were observed in EE2, and combined EE2 and high temperature treatment respectively, which means that the treatment induced feminisation. High temperature induced significant reversal of the phenotypic sex of female (genotypically XX) to male in XX group compared to the control (*p*-value 1.68e-02, **Table 6.10**). A small number of individuals from each group were also tested for sex-linked markers and the genotypes were found to be XX as expected.

A small number of intersex progeny were identified (< 1 %, Total 18 out of 2075 progeny) in all the batches produced, which had a pair of testes containing a few oocytes. Once the XY neo-females were identified for future broodstock, the rest of the normal female (XX) and normal males (XY) from each batch were discarded.

Table 6.10 Survival rate of Nile tilapia from XX♀ × XX♂ cross after 20 days of treatment with EE2 (100 mg/kg) with or without high temperature (28°C or 36°C), and the sex-ratio after three months of rearing at 28°C.

Batch (Hormone dose)	Treatment	Replicate	Initial no. of fish	Final no. of fish	Survival %	*P*-values (G-test), between replicates	*P*-values (G-test), pooled replicate and control	No. of fish sexed	No. of male	No. of female	No. of intersex fish	Male (%)	Sex-reversal, *p*-values (G-test)
1 (100 mg/kg)	28°C	1	45	40	88.89			35	3	32		8.57	
	36°C	1	45	43	95.56		0.81	40	12	28		30	1.68e-02
	EE2	1	45	37	82.22	0.93	0.81	37	1	36		2.7	0.27
		2	45	38	84.44								
	Mean ± SD				**83.33 ± 1.11**								
	EE2 + 36°C	1	45	40	88.89	0.93	0.96	37	0	36	1	0	3.64e-02
		2	45	39	86.67								
	Mean ± SD				**87.78 ± 1.11**								

6.5 Discussion

The results demonstrate that the combined treatment of estrogen hormone and high temperature during the labile period of sex differentiation in Nile tilapia was more efficient at changing the phenotypic sex from male to female compared to the hormone alone in mixed sex, XY and XX groups. Although slightly higher feminisation rates were observed in combined DES and high temperature groups than the combined EE2 and high temperature, the survival rate was higher in EE2 than DES.

Synthetic steroids are more effective than natural steroids at sex reversing fish and the synthetic estrogens, 17α-ethynyloestradiol (EE2) and diethylstilbestrol (DES) are the most potent feminising agents tested in fish so far. The potency of these hormones vary in different fish species and is dependent on the conditions used (Piferrer 2001). For example DES was found to be more efficient than EE2 in feminising *O. aureus* (Rosenstein and Hulata 1994), however the opposite results were found in feminising *O. niloticus* (Gilling *et al*. 1996). Based on the different number of trials and species tested, Piferrer (2001) mentioned that EE2 is about 1.5 times more potent than DES and about three times more potent than E2 (17β-estradiol) in feminising fish.

Steroid hormones have been proven to be effective in sex reversing many different fish species but differences in the efficiency of hormones have been noticed in some species. For example it seems to be much easier to masculinise XX tilapia than to feminise XYs (harder to feminise YYs) while in *Clarias macrocephalus* feminisation is very easy (personal communication David J. Penman). Whereas in Channel catfish it is harder to masculinise XX females than to feminise XY males. In Channel catfish XY female populations can be easily produced using estrogen hormones and in fact paradoxical feminisation (female skewed sex-ratio) has been found in Channel catfish when treated with

testosterone hormone (Goudie *et al*. 1983, 1985). On the other hand exogenous androgens, such as 17α-methyltestosterone, have been found to be effective in the masculinisation of genetically female fish, for example hormonal feed treatment of 17α-methyltestosterone has been extensively and effectively used in commercial scale for producing all male (95 to 100 %) Nile tilapia (Pandian and Sheela 1995; Beardmore *et al*. 2001). However in immersion treatment (single or double) with 17α-methyltestosterone (1000 - 2000 µg/L) at 1 and/or 24 hpf embryo did not show any significant masculinisation in Nile tilapia, on the other hand 11-ketotestosterone at the same dose induced significant masculinisation in Nile tilapia (Gennotte *et al*. 2015). Differences in hormone potency arise from the genetic factors; type, dose and duration of hormone; the different affinities of the hormone receptors, activities of the hormone-receptor complexes, their metabolism and the environment (Devlin and Nagahama 2002).

It is harder to feminise tilapia than to masculinise and very variable feminisation rate has been observed in Nile tilapia when treated with estrogen hormone. Sometimes paradoxical feminisation has been observed when genetically male tilapia fry was treated with high temperature. Therefore in this study the estrogen hormone was combined with high temperature to try to improve the efficacy of feminisation. Results suggested that the combined treatment of estrogen hormone and high temperature induced higher feminisation rate than the hormone alone. One explanation could be, estrogen hormone first activates the expression of estrogen receptors during the gonad differentiation, and the high temperature in the same treatment stimulates the over expression of the receptor. It has been noted that the ovary differentiation starts before testis differentiation in Nile tilapia and the expression of androgen and estrogen receptors is important in changing the sex (Singh 2013; Golan and Levavi-Sivan 2014). Three estrogen receptors (esr1, esr2a, and esr2b) have been found to

express at a higher level at 9 dpf than the androgen receptors in XX and XY progeny (Ijiri *et al*. 2008).

The feminisation rates in the combined treatment of hormone and high temperature were three to six fold higher in mixed sex groups compared to the XY sex groups in batches 2 and 3. About 19 - 46 % of the genetic males (XY) converted their phenotypic sex to female in DES and EE2 with high temperature in the mixed sex groups, whereas only 8 - 18 % females were identified as neo-females in the XY group. However, batch 1 (hormone with temperature) in the XY group showed a high proportion (89 - 94 %) of sex-reversal compared to the mixed sex group. Hormones (DES, EE2) alone during the gonad differentiation stage also showed high sex-reversal in this batch, whereas no sex conversion or only 2 % has been observed in batch 2 and 3 respectively in XY sex group (EE2 only). Combined treatment of estrogen hormone and high temperature improved the rate of feminisation of XY males (and EE was better from the point of view of not causing mortalities) but the results still fluctuated among the batches. This could be due to genetic variation (QTL affecting rates of sex-reversal), or maybe there are some critical parameters that we still haven't standardised (e.g. maybe feeding frequency, or exact start of treatments in relation to developmental rate, which is likely to vary a little among batches).

Gennotte *et al*. (2015) found 9 - 65 % feminisation in XY groups when treated with different doses of EE2 during embryogenesis (1 and/or 24 hpf, single or double 4 hr immersion). Sex inversion by feminising hormone in XY groups could be more effective during the embryonic development. Surprisingly no feminisation was observed in the YY group by estrogen treatments even at higher doses (Gennotte *et al*. 2015). They postulated that the Y-linked sex determinant may prevent the female development pathway when present in two copies or a female development gene present on the X chromosome could be responsible for the absence of susceptibility to feminising hormone in the YY group. They

also mentioned that a higher EE2 concentration is needed for in-feed feminisation of YYs than XYs (150 mg/kg for XY and 500 mg/kg for YY). The follow-up plan (never carried out) from this study was to carry out YY × XY cross to give both types of males followed by testing a range of EE2 concentrations at both 28 and 36^{o}C.

High temperature feminisation has been observed in XY and YY progeny, and high temperature feminisation in XY and YY was efficiently suppressed by combined treatment with high temperature and aromatase inhibitor (AI) (Abucay *et al*. 1999; Kwon *et al*. 2002). They speculated that AI inhibits the activity of aromatase enzyme, which in turn inhibits the production of estrogen from androgen, which causes masculinisation of genetic females and with high temperature in the same treatment triggers the masculinisation process. In the present study no high temperature feminisation was found in XY progeny even though combined treatment of estrogen hormone and high temperature was more efficient than estrogen alone.

In the present study a lower survival rate was found in DES treated batches when compared to the EE2 treated batches and their respective controls. Lower survival rates were also observed at lower concentration of DES (50 and 100 mg/kg) in Nile tilapia (Hamdoon *et al*. 2013). Varadaraj (1989) reported no significant mortality in the Mozambique tilapia even at higher concentration of DES (500 and 1000 mg/kg). Estrogen administration during the labile period of sex differentiation can interfere with the insulin-like growth factor (IGF) system in immune organs leading to an increase in susceptibility to infection which might cause mortality (Shved *et al*. 2009). Shved *et al*. (2007) reported that treating Nile tilapia fry at the dose of 125 mg EE2/kg feed from 10 - 40 dpf during the sensitive period of sexual differentiation induced feminisation of the fish, but also resulted in a severe and persistent reduction in growth in both sexes.

In the present study, high temperature treatment did not induce male biased sex-ratio in the batches produced from XX♀ × XY♂ cross except for batch 4 which showed a significant higher proportion of males (75 %) compared to the control. About 17 % of these males were identified as neo-males (genetically XX) from the genotyping results. Genotyping results also suggested that about 9 % and 13 % males were neo-males from batch 1 and 3 respectively (phenotypic sex-ratio did not deviate from the expected 1:1). In the cross of XX♀ × XX♂, a significant proportion of the genetic females were found to have reversed their sex to male (30 %). The same male was used in cross with a different female and 91 % of the progeny were found to be males, which suggest that the thermosensitivity was of paternal origin in this case (temperature was 0.5^{o}C higher in this case, data unpublished). Temperature treatment of progeny from either thermosensitive males or females showed male skewed sex-ratios, indicating that the sensitivity to temperature is under a genetic basis and is a heritable trait (Baroiller and D'Cotta 2001; Tessema *et al*. 2006; Lühmann *et al*. 2012; Palaiokostas *et al*. 2015).

High temperature induced sex-reversal has been extensively studied in *O. niloticus* and generally induces female to male sex-reversal, which was also found in the present study. High temperature treatment (34 - 36^{o}C) significantly increased the percentage of males in mixed sex groups (69 - 91 %) whereas temperature treatment (≥ 32^{o}C) in all-female groups derived from XX × XX crosses also increased the male proportion from up to 91 % (Baroiller *et al*. 1995).

Response to the high temperature differs between strains/population according to the degree and range of the response. Depending on the strain of Nile tilapia, high temperature (36°C or 37°C) induction either did not change the ovarian differentiation or produced male-skewed sex-ratios (Baroiller *et al*. 1995; Baras *et al*. 2001; Tessema *et al*. 2006; Bezault *et al*. 2007; Azaza *et al*. 2008). Tessema *et al*. (2006) observed 66 % male skewed sex-ratio in

the Lake Manzala population when treated at 36°C whereas in the Lake Rudolph population, same temperature did not show such a high male skewed sex-ratio. Therefore temperature sensitivity of a species is thought to vary between different geographic sources of the species. Conover and Heins (1987) described how in *M. menidia* the magnitude of thermosensitivity decreases with increasing latitude. Female-skewed sex-ratio has been observed during the breeding season when the temperature is relatively low whereas male-skewed sex-ratio has been observed when the temperature is relatively high. Populations of *M. menidia* at different latitudes (different temperature during breeding season) compensate the differences in temperature environments by altering the response of sex-ratio to environmental temperature. It has also been mentioned that thermosensitivity varies between individuals in the same species and a large variation in sex-ratio has been observed in the same strain of Nile tilapia when treated with the same temperature (Azaza *et al*. 2008). It can be concluded that the sensitivity to high temperature varies between individuals, among strains and between populations (Patiño *et al*. 1996; Kwon *et al.* 2002; Baroiller *et al*. 2009a).

Survival rates in thermosensitive batches (36°C) were similar to those of the control group. Others also stated that this temperature does not affect the survival rate of the treated fry (Baras *et al*. 2001; Tessema *et al*. 2006), whereas a slight increase in temperature (to 36.8°C) was observed to increase the masculinisation rate in tilapia but significantly decreased the survival rate (Azaza *et al*. 2008).

6.6 Conclusions

This study first demonstrated the use of combined treatment of estrogen hormone and high temperature in mixed sex, XY and XX groups to feminise genetic males during the critical period of sex differentiation in Nile tilapia. Due to recent advances in mapping sex, the

genotypic sex was verified using tightly sex-linked markers. Although it has been shown that estrogen hormone and high temperature separately have opposite effects on sex-reversal in the gonad differentiation pathway, the combined treatment of estrogen hormone and high temperature proved to produce a higher feminisation rate in genetically male (XY) Nile tilapia, which suggests that while the estrogen hormone is the main activist in the combined treatment in the sex differentiation pathway, high temperature may further stimulate the hormone action. Sex-reversal rates varied among crosses (XX♀ × XY♂, XX♀ × YY♂, and XX♀ × XX♂) and also varied among batches within the cross type, so further studies on sex differentiation and hormonal regulation need to be undertaken in Nile tilapia during the critical period of brain and gonad differentiation, to optimise the treatment and to extend it to the YY genotype.

Acknowledgements

The authors acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance to conduct this research (CSC reference BDCA-2013-5). We thank to Khanlfan-Al Rashdi for producing data from batch 1 (XX♀ × YY♂ cross). We are also thankful to Keith Ranson, Brian Craig, Kirsten McCallum for their help in the Tropical Aquarium and Kerry Bartie for the technical support in the Laboratory at the Institute of Aquaculture, University of Stirling.

Chapter 7

Species identification using SNP markers

Genomic composition of the Molobicus hybrid strain using species diagnostic SNP markers

Bartie K.[1], **Taslima K.**[1,2], Khan M.G.Q.[2], Syaifudin M.[3], Wehner S.[1], Taggart J.B.[1], Bekaert M.[1], de Verdal H.[4,5], Benzie J.[4], McAndrew B.J.[1], Penman D.J.[1]

[1]Institute of Aquaculture, University of Stirling, Stirling FK9 4LA, Scotland, UK
[2]Department of Fisheries Biology and Genetics, Bangladesh Agricultural University, Mymensingh, Bangladesh
[3]Program Study of Aquaculture, Faculty of Agriculture, University of Sriwijaya, South Sumatera, Indonesia
[4]WorldFish Centre, Jalan Batu Maung, 11960 Bayan Lepas, Penang, Malaysia
[5]Cirad, UMR ISEM, 34398 Montpellier Cedex 5, France

Status: To be submitted

Contributions

Experimental design, genomic DNA extraction for GIFT, construction of the ddRAD library for GIFT and sequencing (under the supervision of John B. Taggart), KASP assays, data compilation, allele frequency estimation and statistical analysis, DAPC analysis were conducted by the author of this thesis. This manuscript was first drafted by the author of this thesis who was also entirely involved in the following corrections of the manuscript. The other co-authors contributed in the experimental design, generation production and sample processing, DNA extraction from Molobicus and pure species, KASP assays, DAPC analysis, analysis of generated sequences from MiSeq and editing the manuscript.

7.1 Abstract

Due to the existence of many different tilapia species, the capacity for fertile hybrids, the extensive introduction of tilapia into different countries, their importance in aquaculture and fisheries, and the development of several genetic improvement breeding programmes for tilapia, the identification of tilapia species is of importance for the management of aquaculture and wild stocks. Ten species-diagnostic single nucleotide polymorphic (SNP) markers were developed from a ddRADseq study (four for *Oreochromis mossambicus*, four for *O. niloticus* and two for *O. aureus*) and tested against several putative pure populations, parental strains of the Molobicus hybrid (GIFT from 7th generation and feral *O. mossambicus*), the Molobicus F1 hybrid generation and the seventh Molobicus hybrid generation following selective breeding for higher harvest weight, to identify the genomic status of each species/strain. Putative pure *O. mossambicus*, *O. niloticus* and *O. aureus* from different populations (a total of 75 individuals) were found to be pure based on the 10 species diagnostic SNP markers. The GIFT population (n = 50, generation 19) was found to be composed of mostly *O. niloticus* with small contribution from *O. mossambicus* and no or very little (allele frequency 0.02 for one marker) contribution from *O. aureus* based on 10 SNP markers assayed. The feral *O. mossambicus* were found to be close to pure based on *O. mossambicus* specific allele frequencies of 0.96 – 0.98. Nearly all F1 Molobicus hybrid was found to be heterozygous for the 8 markers, 4 diagnostic for *O. mossambicus* and 4 diagnostic for *O. niloticus*. The *O. niloticus* specific allele (or "not *O. mossambicus* allele") was at a significantly higher frequency for six out of the eight markers in Molobicus hybrid generation 07. Following several generations of selective breeding in favour of the highest harvest weight, Molobicus hybrid (generation 07) had an increasing proportion of *O. niloticus* genome at the expense of *O. mossambicus* genome. This is the first case study with species-diagnostic SNP markers representing the ability to distinguish the three

commercially important tilapia species (*O. niloticus*, *O. mossambicus*, *O. aureus*) and their hybrids, and more markers need to be identified across the whole genome for proper genetic management of the tilapia species and to develop a successful tilapia breeding programme.

Keywords: Tilapia, GIFT, Molobicus, selective breeding, SNP markers, genetic management.

7.2 Introduction

Tilapias belonged to the family Cichlidae, originating from Africa and the Middle East; some of these species are now extensively cultured throughout the world, in approximately 140 countries. Species in the genus *Oreochromis* are commercially the most important and are the biggest contributors to the total world tilapia production. Tilapias were first introduced to Asia (*O. mossambicus* and later *O. niloticus* and *O. aureus)* and have subsequently been widely distributed throughout the world because of their many commercially desirable qualities such as higher growth rates, ability to survive many different aquatic environments and resistance to disease. Following the introduction of tilapias to different countries, it became hard to conserve the genetic quality of the original species and stocks due to poor broodstock management and widespread gene introgression (Taniguchi *et al*. 1985; Eknath *et al*. 1991). Many imports were serial transfers between farms, resulting in low effective population sizes, founder effects and bottlenecks followed by inbreeding depression (Pullin and Capili 1988; Bentsen *et al*. 1998). Hybridisation resulted from the production of various hybrid combinations to generate all-male or nearly all-male fry for farms and poor management, or in some cases misidentification of species, allowing introgression in farmed broodstock. In the Philippines farmed *O. niloticus* were found to be introgressed with *O. mossambicus* (Taniguchi *et al*. 1985). Although

hybridisation has resulted in the loss of purity of some species, hybrids are still being produced intentionally, for example F1 hybrids from *O. niloticus* and *O. aureus* contribute significantly to the total tilapia production in China, the biggest tilapia producer in the world. Although they have lower growth performance than pure *O. niloticus*, the hybrids are preferred due to the high male percentage and better survival in unstable climatic conditions (Thodesen *et al*. 2013).

Introgression and hybridisation are common in wild and feral tilapia species due to anthropogenic influence, leading to severe changes in the genetic structure of native and feral species (Gregg *et al*. 1998; Moralee *et al*. 2000; Adépo-Gourène *et al*. 2006; Angienda *et al*. 2011; Firmat *et al*. 2013; Deines *et al*. 2014).

The maintenance of genetic quality of pure species is very important for successful breeding and stock improvement programmes. Compared to plants and livestock, aquaculture species are far behind in selective breeding programmes to improve commercially important traits (Gjedrem *et al*. 2012). It has been estimated that less than 10 % of the global aquaculture production comes from the genetically improved stocks despite the observed average genetic gain in growth rate of 12.5 % per generation, which is four to five fold higher than that seen in livestock (Quillet *et al*. 2005; Gjedrem *et al*. 2012).

The species makeup of many tilapia populations remains largely unknown. Most of the tilapia breeding programmes have been developed with introduced tilapia species, often with an unknown genetic background. According to Neira (2010) and Rye *et al*. (2010), the highest numbers of breeding programmes (27) are in operation for tilapia around the world and the base populations for at least 10 of the breeding programmes are known to derive from genetically improved farmed tilapia (GIFT), mainly from the widely distributed 5th generation. Therefore it is in need to develop methodology(s) for accurate assessment of the

purity or genetic background of the stocks. Molecular genetic markers offer great potential to distinguish between tilapia species, even sub-species level, and to measure genetic diversity, gene introgression and estimate the genetic parameters (Costa-Pierce 2003). Protein and DNA-based markers have been tested for species identification. Allozyme loci, the protein-based markers have been used to differentiate between tilapia species (Sodsuk and McAndrew 1991). The main obstacles for this technique are that the fish generally need to be killed to obtain the tissue samples, the need to keep the tissue samples frozen at low temperature to preserve enzyme activity and the limited number of detectable, informative markers. Mitochondrial DNA (mtDNA) sequence data has also been used to distinguish tilapia species (Nagl *et al*. 2001; D'Amato *et al*. 2007; Shirak *et al*. 2009) but is of very limited use to analyse hybridisation and introgression as it is in effect a maternally inherited single locus. Other molecular markers such as random amplified polymorphic DNA (RAPD), restriction fragment length polymorphism (RFLP) and microsatellite markers have also been used to tilapia species identification (Agnése *et al*. 1999; Hassanien *et al*. 2004; Deines *et al*. 2014). None of these markers can accurately identify pure tilapia species or estimate introgression. For example, microsatellite markers often display multiple alleles per species with overlapping size ranges.

SNPs are the most frequent type of genetic variation in the genome (Wang *et al*. 1998). SNP markers have wide applicability such as linkage mapping, genome-wide association study, estimation of genetic diversity, population genetic structure, gene introgression, etc. (Van Bers *et al*. 2012; Huang *et al*. 2013; Palaiokostas *et al*. 2013a). Next generation sequencing technologies offer great potential to rapidly generate thousands of reliable SNP markers, which might be associated with specific traits or able to differentiate between species or populations. Reduced representation genome sequencing techniques, such as restriction-site associated DNA sequencing (RADseq), are enable to generate large numbers

of SNP markers, randomly distributed throughout the genome, by sequencing short regions adjacent to the restriction enzyme cut site and are used in genome studies in a wide range of model and non-model organisms (Miller *et al*. 2007; Baird *et al*. 2008; Peterson *et al*. 2012; Toonen *et al*. 2013). These sequencing technologies can be used for population genetics studies for the species with or without any prior sequence data and discover polymorphisms across the population followed by genotyping of a large number of individuals directly from the sequence data (Davey and Blaxter 2010; Hemmer-Hansen *et al*. 2014).

GIFT is one of the most successful fish breeding programmes (started in 1988) and has worldwide impact on aquaculture production (Ponzoni *et al*. 2010). To maintain the broad genetic diversity, the base *O. niloticus* populations for GIFT were sampled from four farmed tilapia stocks (Israel, Singapore, Taiwan and Thailand stocks) and four wild stocks directly imported from Africa (Egypt, Ghana, Kenya and Senegal).

Many *O. niloticus* populations in Asia (from where the base populations were collected for the GIFT development) suffered from poor genetic quality, and were introgressed from undesirable feral *O. mossambicus* (Taniguchi *et al*. 1985; Macaranas *et al*. 1986; Eknath *et al*. 1991). *O. aureus* is native to Jordan River system in Israel and also inhabits in the Nile River system in Egypt. Wild and farmed populations of *O. niloticus* for GIFT development were collected from Egypt and Israel respectively. It was thus possible that *O. niloticus* might have been introgressed with *O. aureus* before collecting the samples for GIFT establishment. Following development of GIFT through selective breeding, they have been distributed in different countries, for example to different countries of Asia, Africa and South America, which may introgress with the existing tilapia species.

Another breeding programme has been conducted in the Philippines (termed as Molobicus) with a view to producing hybrid tilapia that would grow well in higher salinity. The base

populations were GIFT *O. niloticus* from the 7th generation (selected for fast growth) and feral *O. mossambicus* (more salinity tolerant) from the wild in the Philippines. The F1 Molobicus hybrid (cross between *O. niloticus* and *O. mossambicus*) was backcrossed with feral *O. mossambicus* to improve the salinity tolerance, followed by producing generations through selective breeding based on the highest growth performance (de Verdal *et al*. 2014b). Due to the backcross and selection pressure based on harvest weight, it would be interesting to know how the Molobicus genome has been shaped. Therefore the present study was designed to estimate the genomic composition of the parental stocks for Molobicus breeding programme i.e. GIFT (from 19th generation, personal communication John Benzie) and feral *O. mossambicus,* F1 Molobicus hybrid and the 7th generation of Molobicus hybrid strain using SNP markers. GIFT generation 07 was used as parental stock for Molobicus breeding programme, and GIFT generation 19 (few generations later than those used to start the Molobicus programme) was used for this study, and this was the closest generation that we had available to the actual Molobicus parental stock. Putative pure stocks of *O. mossambicus*, *O. niloticus* and *O. aureus* were also tested as positive control. SNP markers for this study were designed based on a ddRADseq study comparing several tilapia species (Syaifudin 2015).

7.3 Materials and methods

7.3.1 Sample collection and preparation

Fin samples from *O. mossambicus* were collected from Stirling (n = 7), South Africa (n = 8) and Singapore (n = 7). The *O. niloticus* samples were collected from Stirling (n = 14) and Ghana (n = 16; 8 from Kpandu and 8 from Nyinuto). Kpandu and Nyinuto are the Lake and River Volta of the Sudano Sahelian Region in Ghana. Fin samples of *O. aureus* were

collected from Stirling (n = 8) and Israel (n = 15). All the populations of *O. mossambicus, O. niloticus* and *O. aureus* were thought to be pure, based on information about origin and the previous ddRADseq study. The GIFT population was produced at the WorldFish Center (Penang, Malaysia), and fin samples of 50 GIFT adult individuals from generation 19 and eight GIFT progeny groups (produced from 4 dams and 4 sires of 50 GIFT individuals tested) were received from WorldFish Center. GIFT progeny were included to the analysis to see the segregation pattern for each marker to the next generation in GIFT. Fin samples of 27 feral *O. mossambicus* (parental stocks for Molobicus breeding programme), 21 F1 hybrid and 58 individuals (from 17 families) from generation 07 (F1 hybrid backcrossed once with *O. mossambicus* followed by seven generations of selective breeding for highest body weight) were received from Philippines. In Molobicus breeding programme, each family was reared in two different farm environments, extensive (n = 24) and intensive (n = 34). The individuals for generation 07 in extensive culture system (n = 24) came from 6 different families (2 – 6 individuals per family) and the individuals (n = 34) from intensive culture system came from 11 different families (2 – 6 individuals per family). Details of the sample information and the sampling points are provided in **Table 7.1** and **Figure 7.1**.

Genomic DNA was extracted using the protein-salt and DNA-isopropanol precipitation method modified from Aljanabi and Martinez (1997). In brief, fin tissue was digested in lysis solution (SSTNE/1 % SDS) and proteinase K (10 mg/mL) at 55^{o}C overnight. Digested samples were treated with RNaseA (2 mg/mL) at 37^{o}C for an hour followed by protein precipitation with 5 M NaCl. DNA was precipitated into absolute isopropanol and dissolved into 5 mM Tris (pH 8.0) until DNA quantification. Quantity and quality of DNA were assessed by Nanodrop spectrophotometer (Labtech International Ltd, UK) and agarose gel electrophoresis.

Table 7.1 Details of the samples: species, origin of the strain/population, type (pure, hybrid or unknown) and number of individuals per strain/population.

Species	**Strain/population**	**Type**	**Number of individuals**
O. mossambicus	Stirling (Zambezi)	Pure	7
	Singapore	Pure	7
	South Africa	Pure	8
O. niloticus	Stirling (Manzala)	Pure	14
	Ghana	Pure	16 (8 – Kpandu, the Lake Volta and 8 - Nyinuto, the River Volta)
O. aureus	Stirling (Manzala)	Pure	8
	Israel	Pure	15
GIFT	Malaysia	?	50
Feral *O. mossambicus*	Philippines	?	27
F1 Hybrid	Philippines	Hybrid	21
Generation 7	Philippines	Hybrid (Extensive culture)	24
Generation 7	Philippines	Hybrid (Intensive culture)	34
GIFT progeny	Malaysia	?	8

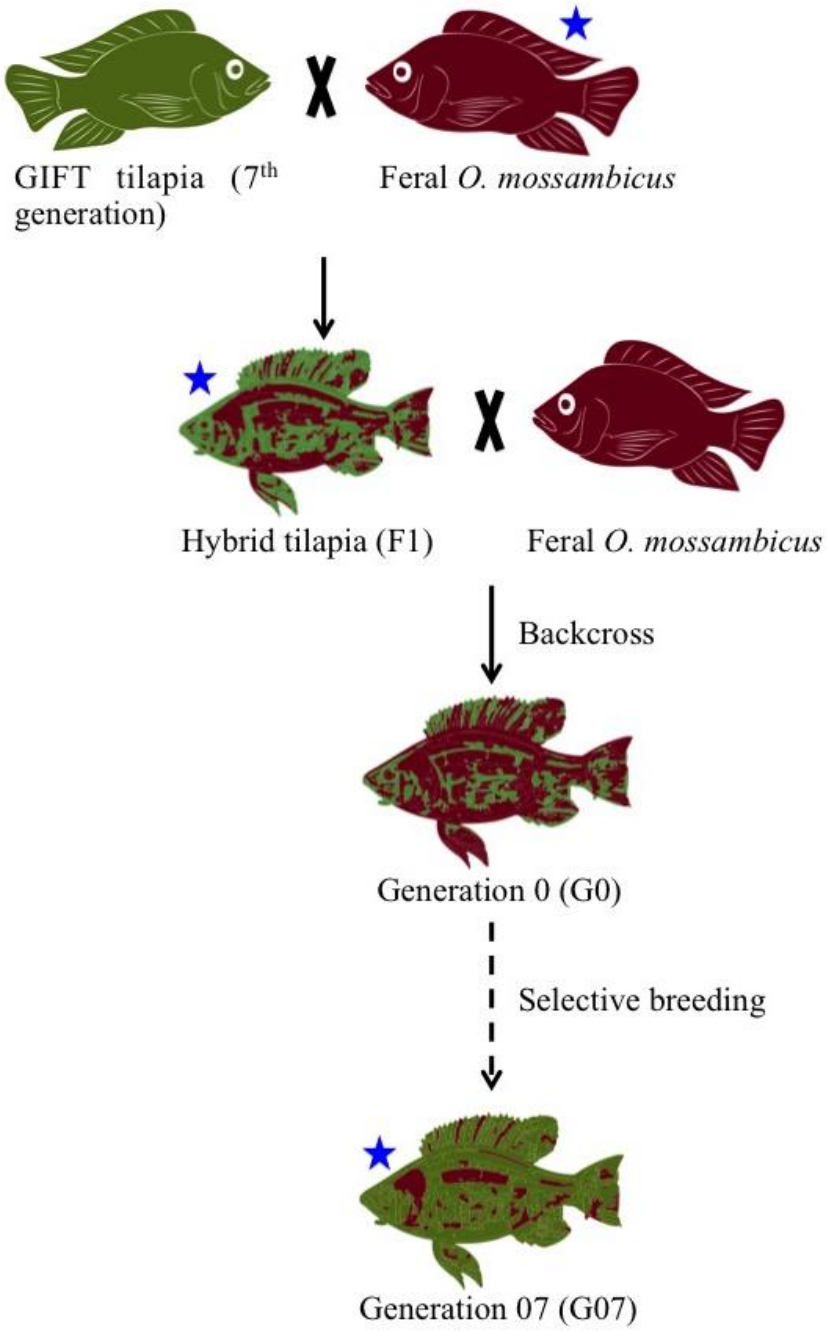


Figure 7.1 Flow diagram of Molobicus breeding programme. An F1 hybrid was produced from GIFT (7th generation) and feral *O. mossambicus*. The F1 hybrid was backcrossed with *O. mossambicus* (genome contribution 75 % from *O. mossambicus* and 25 % from GIFT tilapia). After that 7 generations were produced through selective breeding for higher harvest weight. Blue stars represent the sampling points for this study.

7.3.2 Species-specific diagnostic SNP markers

Species-specific diagnostic SNP markers for different pure species of tilapia were generated from ddRADseq and designed by Syaifudin (2015). A total of 22 SNP markers were identified as candidate species-diagnostic markers for *O. niloticus* (5), *O. mossambicus* (9) and *O. aureus* (8). From the previous study ten of these SNP markers were found to be species diagnostic when tested for pure lines of *O. mossambicus* - 4; *O. niloticus* - 4 and *O. aureus* -2 (unpublished data). Therefore 10 SNP markers were used for this study including another 12 marker for GIFT population. More than 150 bases were retrieved from the region flanking each SNP for designing allele-specific primers. Primers were designed and produced by LGC Genomics (UK) based on the provided sequences. Allelic combinations for each marker with their chromosomal position in the published *O. niloticus* genome and the KASP assay primer sequences are described in **Appendix Table C7.1, C7.2**.

7.3.3 Genotyping of GIFT population using ddRADseq

The SNP data for the GIFT population (n = 50) were derived from the ddRADseq rather than individual KASP assay. Detailed ddRAD library preparation, sequencing and generation of RAD loci are described in Chapter 4, section 4.3.3. Barcode information and the generated paired-end reads for each individual are given in **Appendix Table C7.3**.

Genotype data for each GIFT individual for 22 putative species diagnostic SNP markers (n = 9 for *O. mossambicus*, n = 5 for *O. niloticus*, and n = 8 for *O. aureus*) was extracted from the sequences generated using ddRADseq. The genotype of the GIFT individuals was identified by aligning sequences flanking each SNP against the generated RAD loci from 50 GIFT individuals using Blastn analysis (Blastn 2.2.28+, Altschul *et al*. 1990). The e-value was above 1.00e-50.

7.3.4 SNP marker genotyping using KASP assay

Pure species of *O. niloticus*, *O. mossambicus* and *O. aureus*; feral *O. mossambicus*, Molobicus F1 hybrids, 7th generation Molobicus individuals and GIFT progenies were genotyped for 10 putative species-diagnostic SNP markers out of 22 (4 – *O. mossambicus*, 4 – *O. niloticus*, and 2 – *O. niloticus*). GIFT samples were genotyped for 22 SNP markers (10 putative species-diagnostic markers including another 12 from ddRADseq study of Syaifudin 2015) from ddRAD sequence data and if there was no match for any of the 22 SNP markers to the ddRADseq sequences, these samples were also subjected to KASP genotyping for the particular SNP marker (81 out of 1,100 genotype data were produced from KASP). A few GIFT individuals (n = 5, SNP genotypic data was retrieved from ddRADseq) was also genotyped with SNP marker genotyping system to verify the data from both ddRADseq and marker genotyping system. Individuals were genotyped using fluorescence-based Kompetitive Allele Specific end-point PCR (KASP) genotyping system (LGC Genomics, UK). A 5 µL PCR reaction volume was prepared with c. 25 ng DNA and the PCR cyclic conditions were the initial denaturation at 94°C for 15 min followed by 10 touchdown cycles (94°C for 20 sec and touchdown 65°C for 1 min, -0.8°C per cycle), and 34 cycles of amplification at 94°C for 20 sec and 57°C for 1 min. Fluorescence was then detected at ambient temperature using Techne Quantica® machine (Barloworld Scientific Ltd UK) and individual was genotyped based on the fluorescence signal. An allelic discrimination analysis was performed to determine the single nucleotide differences using the inbuilt Quansoft software (version 1.1.21).

7.3.5 Statistical analysis

Allele frequencies and the deviation of allele frequency (Chi-square goodness of fit test) from the expected ratio (1:3) were calculated using R (version 3.1.3). Discriminant analysis of principal components (DAPC), a multivariate analysis designed to identify the genetically related individuals followed by forming clusters, was performed using the R package "adegenet".

Data access

The raw sequence data for this study were submitted to the EBI's European Nucleotide Archive (ENA) Sequence Read Archive (SRA), study accession number PRJEB13792.

7.4 Results

7.4.1 Reference (pure) species of tilapia

The *O. mossambicus* samples from three populations showed only the allele proposed to be specific to this species for each of the four putative *O. mossambicus* diagnostic markers (homozygote and the allele frequency was 1.00). Allele proposed to be diagnostic for *O. niloticus* (4 markers) and *O. aureus* (2 markers) was not detected in any of the *O. mossambicus* populations analysed (**Table 7.2**).

Table 7.2 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for the *O. mossambicus* individuals in three different populations. These include population, species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker.

Population	Species	Species-diagnostic marker	**Number of individual per genotype**			**Allele frequency**	
			Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
Stirling	*O. mossambicus*	*Omos2007*	7	0	0	1	0
		Omos2657	7	0	0	1	0
		Omos3481	7	0	0	1	0
		Omos7956	7	0	0	1	0
	O. niloticus	*Onil2675*	0	0	7	0	1
		Onil3057	0	0	7	0	1
		Onil5782	0	0	7	0	1
		Onil9497	0	0	7	0	1
	O. aureus	*Oaur966*	0	0	7	0	1
		Oaur9418	0	0	7	0	1
South Africa	*O. mossambicus*	*Omos2007*	8	0	0	1	0
		Omos2657	8	0	0	1	0
		Omos3481	8	0	0	1	0
		Omos7956	8	0	0	1	0
	O. niloticus	*Onil2675*	0	0	8	0	1
		Onil3057	0	0	8	0	1
		Onil5782	0	0	8	0	1
		Onil9497	0	0	8	0	1
	O. aureus	*Oaur966*	0	0	8	0	1
		Oaur9418	0	0	8	0	1
Singapore	*O. mossambicus*	*Omos2007*	7	0	0	1	0
		Omos2657	7	0	0	1	0
		Omos3481	7	0	0	1	0
		Omos7956	7	0	0	1	0
	O. niloticus	*Onil2675*	0	0	7	0	1
		Onil3057	0	0	7	0	1
		Onil5782	0	0	7	0	1
		Onil9497	0	0	7	0	1
	O. aureus	*Oaur966*	0	0	7	0	1
		Oaur9418	0	0	7	0	1

In the case of the *O. niloticus* reference population from Ghana, all four putative *O. niloticus* specific markers were found to only show the allele proposed to be specific for this species (homozygote and the allele frequency was 1.00). In the Stirling *O. niloticus* population, these loci also showed only the putative *O. niloticus* specific alleles except for the locus *Onil2675* (**Table 7.3**). Two individuals (out of the 14 assayed) were found to be heterozygous for that marker. None of the allele specific for *O. mossambicus* (4 markers) and *O. aureus* (2 markers) was observed in the *O. niloticus* samples analysed (**Table 7.3**).

Table 7.3 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for the *O. niloticus* individuals in two different populations. These include population, species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker.

			Number of individual per genotype			**Allele frequency**	
Population	Species	Species-diagnostic marker	Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
Stirling	*O. mossambicus*	*Omos2007*	0	0	14	0	1
		Omos2657	0	0	14	0	1
		Omos3481	0	0	14	0	1
		Omos7956	0	0	14	0	1
	O. niloticus	*Onil2675*	12	2	0	0.93	0.07
		Onil3057	14	0	0	1	0
		Onil5782	14	0	0	1	0
		Onil9497	14	0	0	1	0
	O. aureus	*Oaur966*	0	0	14	0	1
		Oaur9418	0	0	14	0	1
Ghana	*O. mossambicus*	*Omos2007*	0	0	16	0	1
		Omos2657	0	0	16	0	1
		Omos3481	0	0	16	0	1
		Omos7956	0	0	16	0	1
	O. niloticus	*Onil2675*	16	0	0	1	0
		Onil3057	16	0	0	1	0
		Onil5782	16	0	0	1	0
		Onil9497	16	0	0	1	0
	O. aureus	*Oaur966*	0	0	16	0	1
		Oaur9418	0	0	16	0	1

In case of *O. aureus* populations, individuals were found to be pure and had allele only specific to *O. aureus* based on the two markers tested (homozygote and the allele frequency was 1.00). Allele-specific to *O. mossambicus* (4 markers) or *O. niloticus* (4 markers) was not detected in any of the *O. aureus* populations studied (**Table 7.4**).

Table 7.4 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for the *O. aureus* individuals in two different populations. These include population, species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker.

Population	Species	Species-diagnostic marker	**Number of individual per genotype**			**Allele frequency**	
			Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
Stirling	*O. mossambicus*	*Omos2007*	0	0	8	0	1
		Omos2657	0	0	8	0	1
		Omos3481	0	0	8	0	1
		Omos7956	0	0	8	0	1
	O. niloticus	*Onil2675*	0	0	8	0	1
		Onil3057	0	0	8	0	1
		Onil5782	0	0	8	0	1
		Onil9497	0	0	8	0	1
	O. aureus	*Oaur966*	8	0	0	1	0
		Oaur9418	8	0	0	1	0
Israel	*O. mossambicus*	*Omos2007*	0	0	15	0	1
		Omos2657	0	0	15	0	1
		Omos3481	0	0	15	0	1
		Omos7956	0	0	15	0	1
	O. niloticus	*Onil2675*	0	0	15	0	1
		Onil3057	0	0	15	0	1
		Onil5782	0	0	15	0	1
		Onil9497	0	0	15	0	1
	O. aureus	*Oaur966*	15	0	0	1	0
		Oaur9418	15	0	0	1	0

7.4.2 GIFT tilapia population

The GIFT samples (50 adult individuals from generation 19) were first screened for 22 putative species-diagnostic SNP markers (the 10 putative species-diagnostic SNPs that were screened on the samples of the three species described above, plus another 12 putative species-diagnostic SNPs identified only from the ddRADseq study of Syaifudin 2015; in

total nine *O. mossambicus*-specific, five *O. niloticus*-specific and eight *O. aureus*-specific markers). GIFT tilapia was found to be composed of mostly *O. niloticus* with the varying proportion of *O. mossambicus* and *O. aureus* based on 22 putatively diagnostic markers. Only one marker (putatively diagnostic for *O. niloticus*) indicated that the GIFT samples were pure *O. niloticus*, and six (putatively diagnostic for *O. mossambicus*) and two (putatively diagnostic for *O. aureus*) markers suggested that there was no contribution of alleles specific for *O. mossambicus* and *O. aureus* respectively (**Table 7.5**). When the markers were reduced to only the 10 more thoroughly validated species-diagnostic SNPs (out of 22), GIFT was found to be composed of mostly *O. niloticus* with very little to high contribution from *O. mossambicus* (allele frequency 0.04 to 0.38 for 4 markers, 2 diagnostic for *O. niloticus* and 2 for *O. mossambicus*) and zero to very little contribution from *O. aureus* (allele frequency 0.02 for one marker).

A majority of the markers (5 out of 8, the most thoroughly tested *O. niloticus* - 4 and *O. mossambicus* – 4 markers) showed > 90 % alleles putatively diagnostic for *O. niloticus* or not *O. mossambicus* or not *O. aureus* in GIFT (**Table 7.5**, **7.6**). When 2 markers diagnostic for *O. aureus* were included, seven out of ten markers also showed > 90 % alleles putatively diagnostic for *O. niloticus* in GIFT (**Table 7.5**, **7.6**; **Figure 7.2**). When only *O. niloticus* (n = 5) and *O. mossambicus* (n = 9) were considered, a majority of the markers (10 out of 14) still showed the same results. When all the markers were considered, 16 out of 22 showed similar results. A smaller proportion of the marker (ranged from three out of eight up to four out of 22) had 60 - < 80 % alleles putatively diagnostic for *O. niloticus* or not *O. mossambicus* or *O. aureus* in GIFT and an even smaller proportion of marker (2 out of 22, only in the larger dataset) showed 20 – 40 % alleles putatively diagnostic for *O. niloticus* or not *O. mossambicus* (n = 1) or not *O. aureus* (n = 1) in GIFT (**Table 7.5**, **7.6)**.

Table 7.5 Genotype and allele frequencies for the 22 putative species-diagnostic SNP markers (n = 9 for *O. mossambicus*, n = 5 for *O. niloticus*, and n = 8 for *O. aureus*) for the GIFT population (n = 50). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker.

		Number of individual per genotype			**Allele frequency**	
Species	Species-diagnostic marker	Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
O. mossambicus	*Omos2007*	4	15	31	0.23	0.77
	Omos2657	0	4	46	0.04	0.96
	Omos3481	0	0	50	0.00	1.00
	Omos7956	0	0	50	0.00	1.00
	Omos10120	0	0	50	0.00	1.00
	Omos10818	37	0	26	0.74	0.26
	Omos3582	0	0	50	0.00	1.00
	Omos8084	0	0	50	0.00	1.00
	Omos4092	0	0	50	0.00	1.00
O. niloticus	*Onil2675*	20	22	8	0.62	0.38
	Onil3057	29	15	6	0.73	0.27
	Onil5782	50	0	0	1.00	0.00
	Onil9497	40	10	0	0.90	0.10
	Onil1276	42	8	0	0.92	0.08
O. aureus	*Oaur966*	1	0	49	0.02	0.98
	Oaur9418	0	0	50	0.00	1.00
	Oaur8029	1	8	41	0.10	0.90
	Oaur3001	6	13	31	0.25	0.75
	Oaur2890	0	0	50	0.00	1.00
	Oaur3873	0	2	48	0.02	0.98
	Oaur5416	33	2	15	0.68	0.32
	Oaur4411	0	8	42	0.08	0.92

Table 7.6 Frequency distribution table including the percent match to *O. niloticus* (or not *O. mossambicus* or *O. aureus*) for putative species-diagnostic 8 (thoroughly tested 4 for *O. mossambicus*, 4 for *O. niloticus*), 10 (4 for *O. mossambicus*, 4 for *O. niloticus* including 2 for *O. aureus*), 14 (9 for *O. mossambicus* and 5 for *O. niloticus*) and 22 (9 for *O. mossambicus*, 5 for *O. niloticus* and 8 for *O. aureus*) markers for GIFT population.

% matches to *O. niloticus*	**8 markers**	**10 markers**	**14 markers**	**22 markers**
0 - <10				
10 - <20				
20 - <30			1	1
30 - <40				1
40 - <50				
50 - <60				
60 - <70	1	1	1	1
70 - <80	2	2	2	3
80 - <90				
90 - <100	2	3	3	7
100	3	4	7	9

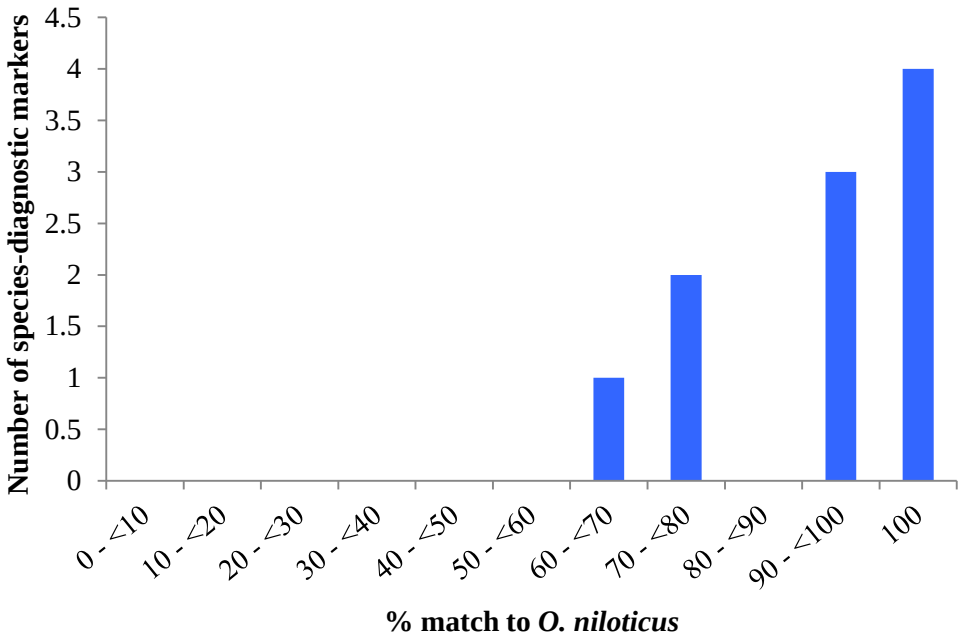


Figure 7.2 Frequency distribution graph including percent match to the *O. niloticus* for the 10 putative species-diagnostic markers (4 for *O. mossambicus*, 4 for *O. niloticus* and 2 for *O. aureus*).

Eight progeny groups (two each from four single pair crosses) were also tested for the 10 SNP markers. Only four markers (out of ten; 2 for *O. mossambicus*, 1 for *O. niloticus* and 1 for *O. aureus*) suggested that GIFT population composed of *O. niloticus* whereas in GIFT progeny six out of ten markers (3 for *O. mossambicus*, 1 for *O. niloticus* and 2 for *O. aureus*) suggested that GIFT progeny composed of *O. niloticus* with no contribution from *O. mossambicus* or *O. aureus*. For the rest of the markers the frequency of the allele diagnostic to *O. niloticus* or not *O. mossambicus* or not *O. aureus* increased in the GIFT progeny (ranged from 0.02 to 0.27) compared to the GIFT population except *Onil9497* (frequency of the allele diagnostic to *O. niloticus* reduced by 0.02).

Table 7.7 Genotype and allele frequencies for the 10 putative species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for the GIFT progeny (n = 8). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker.

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency	
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
O. mossambicus	*Omos2007*	0	8	0	0.50	0.50
	Omos2657	0	0	8	0.00	1.00
	Omos3481	0	0	8	0.00	1.00
	Omos7956	0	0	8	0.00	1.00
O. niloticus	*Onil2675*	3	5	0	0.69	0.31
	Onil3057	7	1	0	0.94	0.06
	Onil5782	8	0	0	1.00	0.00
	Onil9497	6	2	0	0.88	0.12
O. aureus	*Oaur966*	0	0	8	0.00	1.00
	Oaur9418	0	0	8	0.00	1.00

7.4.3 Feral *O. mossambicus* tilapia (parent of Molobicus hybrid)

Nearly all individuals were homozygous for the allele (*Omos2007*) specific for *O. mossambicus* (allele frequency was 0.98). For the rest of the *O. mossambicus* specific markers (*Omos2657*, *Omos3481*, *Omos7956*), frequency for the allele specific to *O. mossambicus* was 0.96. In case of the *O. niloticus* specific markers, the frequency for the allele specific to *O. niloticus* was 0.02 for the two markers (*Onil2675*, *Onil3057*), 0.04 for the marker *Onil9497* and 0.06 for the rest of the marker (*Onil5782*). There was no contribution of *O. aureus* specific allele to the feral *O. mossambicus* populations studied (**Table 7.8**). The number of individuals per marker varied from 24 to 27 due to the SNP assay failure.

Table 7.8 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for feral *O. mossambicus* population. These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker for feral *O. mossambicus* population (n = 24-27).

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency	
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
O. mossambicus	*Omos2007*	26	1	0	0.98	0.02
	Omos2657	24	2	0	0.96	0.04
	Omos3481	23	2	0	0.96	0.04
	Omos7956	23	2	0	0.96	0.04
O. niloticus	*Onil2675*	0	1	24	0.02	0.98
	Onil3057	0	1	24	0.02	0.98
	Onil5782	0	3	21	0.06	0.94
	Onil9497	0	2	24	0.04	0.96
O. aureus	*Oaur966*	0	0	27	0.00	1.00
	Oaur9418	0	0	27	0.00	1.00

7.4.4 F1 Molobicus hybrid (GIFT × feral *O. mossambicus*)

F1 Molobicus hybrid was produced from crossing GIFT (7^{th} generation) and feral *O. mossambicus.* The data (**Table 7.9**) showed that the F1 fish were nearly all heterozygotes for the markers diagnostic for *O. niloticus* and *O. mossambicus*. Only one marker (*Omos2007*) from the *O. mossambicus* diagnostic markers showed slight deviation (allele frequency 0.45:0.55) from the expected ratio. Based on the *O. niloticus* specific markers, only marker *Onil3057* showed a large deviation from this (five homozygotes). None of the individuals showed *O. aureus*-specific alleles for markers *Oaur966* or *Oaur9418* (**Table 7.9**). The number of individuals per marker varied from 20 to 21 due to the SNP assay failure.

Table 7.9 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for F1 Molobicus hybrid (GIFT and *O. mossambicus*). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker for F1 hybrid (n = 20-21).

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency	
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele
O. mossambicus	*Omos2007*	2	19	0	0.55	0.45
	Omos2657	0	20	0	0.50	0.50
	Omos3481	0	21	0	0.50	0.50
	Omos7956	0	20	0	0.50	0.50
O. niloticus	*Onil2675*	1	20	0	0.52	0.48
	Onil3057	0	15	5	0.38	0.62
	Onil5782	0	21	0	0.50	0.50
	Onil9497	1	20	0	0.52	0.48
O. aureus	*Oaur966*	0	0	21	0.00	1.00
	Oaur9418	0	0	21	0.00	1.00

7.4.5 7th generation of Molobicus hybrid tilapia

The F1 hybrid was backcrossed once with feral *O. mossambicus* and seven generations from the backcross were subsequently produced through selective breeding based on selection for higher harvest weight. Significant increase in the frequency of the *O. niloticus* allele for the three *O. niloticus* specific markers or significant decrease in the case of the *O. mossambicus* specific allele for the three *O. mossambicus* specific markers were observed for the generation 07 (**Table 7.10**), which means through several generations of selective breeding the proportion of the *O. niloticus* alleles was significantly increased at the expense of the *O. mossambicus* alleles. Allele specific to *O. aureus* was not detected in any of the individuals in Molobicus generation 07 studied with the two relevant markers.

Table 7.10 Genotype and allele frequencies for the ten selected species-diagnostic SNP markers (n = 4 for *O. mossambicus*, n = 4 for *O. niloticus*, and n = 2 for *O. aureus*) for Molobicus generation 07 (n = 58). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker and Chi-square *p*-value.

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency		*P*-value
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele	
O. mossambicus	*Omos2007*	38	14	6	0.78	0.22	0.520
	Omos2657	24	25	9	0.63	0.37	0.003
	Omos3481	15	24	19	0.47	0.53	0.010
	Omos7956	23	19	16	0.56	0.44	0.011
O. niloticus	*Onil2675*	9	25	24	0.37	0.63	0.003
	Onil3057	3	29	26	0.30	0.70	0.198
	Onil5782	20	25	13	0.56	0.44	0.011
	Onil9497	7	28	23	0.36	0.64	0.005
O. aureus	*Oaur966*	0	0	58	0.00	1.00	
	Oaur9418	0	0	58	0.00	1.00	

Generation 07 was splited into two different environments such as extensive and intensive. In the extensive culture system, only 3 loci (out of 8); two for the *O. mossambicus* specific markers (*Omos3481*, *Omos7956*) and one for *O. niloticus* marker (*Onil5782*) showed significant increase in the frequency of the *O. niloticus* allele or significant decrease in the case of the *O. mossambicus* specific allele (**Table 7.11**). On the other hand all the markers showed significant deviation of allele frequency from the expected ratio in the intensive culture system except one (*Omos2007*, **Table 7.12**).

Table 7.11 Genotype and allele frequencies for the eight selected species-diagnostic SNP markers (n = 4 for *O. mossambicus* and n = 4 for *O. niloticus*) for Molobicus generation 07 reared in extensive culture system (n = 24). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker and Chi-square *p*-value.

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency		*P*-value
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele	
O. mossambicus	*Omos2007*	14	8	2	0.75	0.25	1.000
	Omos2657	16	6	2	0.79	0.21	0.505
	Omos3481	7	11	6	0.52	0.48	0.010
	Omos7956	10	8	6	0.58	0.42	0.008
O. niloticus	*Onil2675*	2	6	16	0.21	0.79	0.505
	Onil3057	2	6	16	0.21	0.79	0.505
	Onil5782	7	13	4	0.56	0.44	0.011
	Onil9497	3	10	11	0.33	0.77	0.182

Table 7.12 Genotype and allele frequencies for the eight selected species-diagnostic SNP markers (n = 4 for *O. mossambicus* and n = 4 for *O. niloticus*) for Molobicus generation 07 reared in intensive culture system (n = 34). These include species, putative species-diagnostic marker, number of individuals per genotype followed by allele frequency for each marker and Chi-square *p*-value.

Species	Species-diagnostic marker	Number of individual per genotype			Allele frequency		*P*-value
		Homozygous for species-diagnostic allele	Heterozygote	Homozygous for alternate allele	Species-diagnostic allele	Alternate allele	
O. mossambicus	*Omos2007*	24	6	4	0.79	0.21	0.401
	Omos2657	8	19	7	0.51	0.49	0.011
	Omos3481	8	13	13	0.43	0.57	0.014
	Omos7956	13	11	10	0.54	0.46	0.013
O. niloticus	*Onil2675*	7	19	8	0.48	0.52	0.011
	Onil3057	1	23	10	0.37	0.63	0.025
	Onil5782	13	12	9	0.56	0.44	0.014
	Onil9497	4	18	12	0.38	0.62	0.012

Based on the DAPC analysis (based on 8 SNP markers, 4 – *O. mossambicus*, 4 – *O. niloticus*), pure species of *O. mossambicus* and *O. niloticus* were divided into two distinct groups using components 1 and 2 (**Figure 7.3a**) or components 1 and 3 (**Figure 7.3b**). *O. mossambicus* from different populations formed only one cluster whereas *O. niloticus* from Stirling and Ghana were very closely related but formed two separate clusters. The GIFT population was highly variable and closer to *O. niloticus* than to *O. mossambicus*. The GIFT progeny formed a cluster slightly separated from the GIFT broodstock. Although feral *O. mossambicus* showed slight sequence diversity but was found to be closer to pure *O. mossambicus* (**Figure 7.3**). The F1 hybrid formed one cluster between the pure *O. niloticus* and *O. mossambicus*. Generation 07 showed high diversity and formed a cluster slightly away from the feral *O. mossambicus*, overlapping the F1 hybrids. Generation 07 reared in extensive culture system was closer to the feral *O. mossambicus* compared to the group reared in the intensive culture system (**Figure 7.3**).

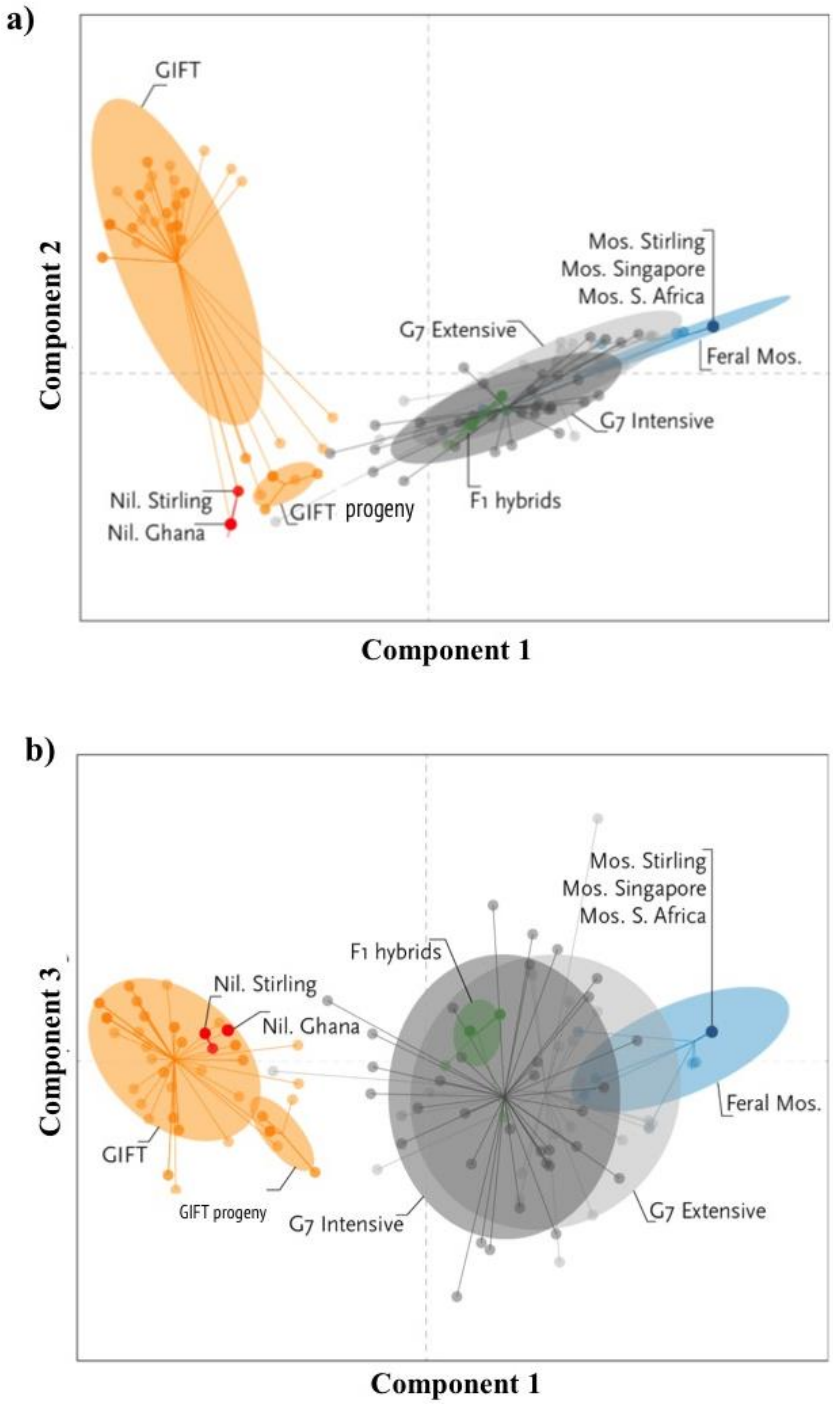


Figure 7.3 Discriminant analysis of principal components; a) component 1 and component 2, b) component 1 and component 3. Population/groups are indicated in different colours; Mos. - *O. mossambicus*, Nil. - *O. niloticus*.

7.5 Discussion

The aim of this research was to validate and use species-specific SNP markers to investigate the genomic contribution of the parental species to GIFT, the F1 hybrid between GIFT and feral *O. mossambicus* and to the Molobicus hybrid line after seven generations of selective breeding. This was done primarily using a single-plex SNP genotyping platform (GIFT data were generated from ddRADseq) and different populations of pure species were used as reference populations.

All the reference populations appeared to be pure based on the 10 species-specific diagnostic SNP markers regardless of the populations studied. Those populations originated from the different sources. Stirling *O. niloticus* and *O. aureus* were collected from the Lake Manzala, Egypt in 1979 (McAndrew *et al*. 1988) whereas Stirling *O. mossambicus* was collected from the Zambezi River, Zimbabwe in 1985 (Majumdar and McAndrew 1986). *O. niloticus* had been introduced into Zambia around 1960s for cage culture in Lake Kariba and for farming in the Kafue River catchment; both are located in the middle of Zambezi River system. It has been reported that the *O. niloticus* has escaped from those systems to Zimbabwe and hybridised with indigenous *O. mossambicus* (Canonico *et al*. 2005), in our study, there was no evidence of any mixing. In case of Stirling *O. niloticus* samples analysed, two copies of an alternate allele found at one marker (*Onil2675*). The explanation could be either the stock is pure but there is a rare alternate allele in the Stirling stock, so the marker is not really species-diagnostic or the marker is species-diagnostic and the present analysis has picked up trace introgression. Whereas the *O. niloticus* population from Ghana was found to be pure. *O. mossambicus* population was also collected from Natal-South Africa (native species) and Singapore. *O. niloticus* was extensively introduced throughout Africa including SW South Africa and Natal (Skelton 1993). There is an

evidence of hybridisation between introduced *O. niloticus* and native *O. mossambicus* in the Limpopo River of South Africa (Van der Waal and Bills 2000). Another study reported that *O. mossambicus* population from River and Farm including Natal and Limpopo River was found as pure *O. mossambicus* based on mtDNA marker (D'Amato *et al*. 2007). Whereas *O. niloticus* collected from Limpopo River was found to be introgressed with *O. mossambicus* in the same study. In our study based on SNP markers the *O. mossambicus* population from Natal-South Africa and Singapore was found to be pure. The *O. aureus* population from Stirling and Israel was also found to be pure and there was no evidence of introgression from other tilapia species.

The GIFT individuals (from generation 19) were genotyped using ddRADseq data for 22 putative species-diagnostic SNP markers (9 – *O. mossambicus*, 5 – *O. niloticus* and 8 – *O. aureus*) to get the possible species mix resulting from the original founder groups. SNPs for this study were selected from 12 different linkage groups. As ddRADseq is a reduced representation genome sequencing technique which sequences only the region flanking a restriction enzyme cut site, not all the genotype data for 22 SNPs markers could be retrieved for all 50 GIFT individuals. Therefore about 7.36 % of the genotype data for GIFT (81 genotype data out of 1,100) were generated using the KASP genotyping system. A few samples (n = 5) also verified with both ddRADseq and KASP genotyping system which gave consistent genotype results. A majority of the markers (7 out of 10) suggested that the GIFT (generation 19) was mostly composed of *O. niloticus* with no to very low (frequency ranged from 0.02 – 0.1) contribution from *O. mossambicus* or *O. aureus* and in a larger dataset, 16 out of 22 markers also showed the same results. Two markers (out of 22; *Omos10818* and *Oaur5416*) suggested that GIFT had a very high proportion of *O. mossambicus* or *O. aureus* specific allele (frequency 0.74 and 0.68 respectively). GIFT is a synthetic mixture of pure and crossbreed groups of wild and farmed *O. niloticus* strains

(Ponzoni *et al*. 2010) which might have already been introgressed with other tilapia species before sample collection for GIFT development. Early problems with species identification between *O. aureus* and *O. niloticus*, and the widespread use of interspecific hybridisation in the 1970 and 80s to generate all-male or nearly all-male fry resulted in widespread introgression between *O. aureus* and *O. niloticus* in many commercial populations. The actual contribution of possibly mixed commercial strains and pure wild collections is not recorded. *O. aureus* is native to Israel and the Nile River system in Egypt. The base population of *O. niloticus* for GIFT development was assembled from Egypt and Israel, so there is the chance of gene introgression between these species (Zak *et al*. 2014). It has been reported that the *O. niloticus* production in Asia was significantly reduced due to poor genetic management and gene introgression from undesirable feral *O. mossambicus* and this was one of the main reasons for initiating the GIFT project (Macaranas *et al*. 1986; Taniguchi *et al*. 1985; Eknath *et al*. 1991). The farmed populations of *O. niloticus* were collected from Asia for GIFT development, which might be the reason of contributing *O. mossambicus* or *O. aureus* alleles in GIFT. The results using the 10 highly diagnostic SNP markers, the frequency of the *O. mossambicus* or *O. aureus*-diagnostic alleles in the GIFT composition reduced significantly; and those 10 SNP markers proved to be diagnostic for putative pure species and also gave better picture of GIFT composition. DAPC analysis also showed that GIFT formed a separate cluster (based on component 1 and 2) that was closer to pure *O. niloticus* than *O. mossambicus*. A small number of progeny (their parents were included in the 50 GIFT population analysed) were also tested for 10 SNP markers in this study. The frequency of the allele for all the markers putatively diagnostic for *O. niloticus* has increased except for the marker *Onil9497* (frequency reduced by 0.02) and the frequency of the allele diagnostic for the *O. mossambicus* or *O. aureus* has decreased in the GIFT progeny. Those progeny were produced through selective breeding for higher harvest

rate and progenies were also found to be much closer to *O. niloticus* than the GIFT individuals in DAPC analysis.

Xia *et al*. (2014) used genotype data from 101 SNP markers from LG23 to compare GIFT with a *O. niloticus* stock held in Shanghai, China (selected line of *O. niloticus* for growth traits, that originated from Egypt), and found that they clustered together with the *O. niloticus* (individual-based Neighbor Joining phylogenetic tree). Later on Xia *et al*. (2015) extended their research using whole genome sequencing of different populations of GIFT (Singapore, Shanghai and Guangzhou; generation was not mentioned) and found that GIFT was closely related to the Shanghai *O. niloticus* but formed a separate cluster (the reasons could be the GIFT and Shanghai *O. niloticus* are two different tilapia lines and selection was done in different environments).

Perhaps surprisingly, feral *O. mossambicus* from the Philippines were found to be nearly pure, with a very small contribution from *O. niloticus* (allele frequency < 0.06) and no *O. aureus* contribution. Although this is an introduced species, it seems to have maintained its genetic purity in the wild. *O. mossambicus* was introduced into the lakes and reservoirs in Philippines since the mid 1950s and *O. niloticus* was introduced later and is farmed extensively there, so there is the possibility of having gene flow from *O. niloticus* to *O. mossambicus* through escapes from the farms and subsequent hybridisation and gene introgression (Pullin *et al*. 1997). However, no evidence for large-scale introgression was found in the present study.

All the Molobicus F1 hybrid individuals were found to be heterozygote for *O. niloticus* (n = 4) and *O. mossambicus* (n = 4) putatively diagnostic markers except 1 – 5 individuals with no contribution from *O. aureus* diagnostic allele. As the F1 Molobicus hybrid was produced by crossing GIFT (from 7th generation) and feral *O. mossambicus*, and none of the parental

lines (although GIFT generation 19 was analysed for this study which was few generations later than used for Molobicus breeding programme) were found to be 100 % pure based on the SNP markers studied, some deviation of the genotypes from the expected was detected in the F1 hybrid population. The feral *O. mossambicus* were very close to 100 % frequency of the *O. mossambicus* (or not *O. niloticus*) allele for all eight relevant loci, but the GIFT varied considerably from 100 % for three markers. It thus doesn't seem justified to expect 100 % heterozygotes in all eight of these markers in the F1 on this basis (even though that has been observed apart from marker *Onil3057*, where the excess of the *O. niloticus* allele was in the opposite direction to that expected from the data for this marker in the GIFT population). There is a question whether the F1 sample size (n = 20-21) is enough to justify this or could the actual GIFT generation that was the parents of the Molobicus F1 hybrid have had different allele frequencies?

After seven generations of selection for highest harvest weight, it seems like the allele frequency moved significantly from the original composition of the strain before selection (25 % *O. niloticus* and 75 % *O. mossambicus*) towards the *O. niloticus* (or not *O. mossambicus* or *O. aureus*) in the Molobicus strain for all the markers tested except two (in combined data set from extensive and intensive culture systems). The differences in allele frequency were obvious between the extensive and intensive culture environments. Only three markers (out of 8) showed significant changes in allele frequency (compared to 3:1) in extensive culture system and the individuals also clustered towards the feral *O. mossambicus* in DAPC analysis. However in the intensive culture system, all the markers showed significant allele frequency shift towards *O. niloticus* except one (*Omos2007*). Based on *Omos2007* marker, feral *O. mossambicus* (actual parent of Molobicus breeding programme) was found to be close to pure *O. mossambicus* (allele frequency was 0.98); whereas allele putatively diagnostic to *O. mossambicus* for the marker *Omos2007* was

present in slightly high proportion (frequency was 0.23) in GIFT (parent in the Molobicus breeding programme; although the GIFT generation used in this study was more than 10 generations later than the generation used in Molobicus breeding programme). In Molobicus breeding programme, each family was reared in two distinct environments i.e. extensive and intensive, and the broodtsock from extensive and intensive culture systems were kept separate to produce the next generation, which suggest that the families from extensive and intensive are from two separate lines. In intensive culture system, fish were maintained at high stocking density and fed *ad libitum* with the salinity 21.6 ± 7.66 ppt whereas the fishes in extensive culture system were maintained at low stocking density with no external feed input and the salinity was 14.9 ± 6.93 ppt. The salinity seems to be different in extensive and intensive culture system; and it has been found that the optimum salinity for the *O. niloticus* ranged from 0 to 10 ppt (Villegas 1990), and they can grow well at 18 ppt in direct transfer and 36 ppt in gradual transfer (Al-Amoudi 1987; Whitefield and Blaber 1979). In the de Verdal *et al*. (2014b) study, average body weight of the Molobicus strain was higher in the intensive culture system than the extensive system; and 65.8 and 38.7 % greater body weight was obtained in intensive culture system for male and female respectively. After four generations of selection, average body weight of the Molobicus strain was increased by 50 g (12.5 g per generation). Fish were well fed with the controlled environment in the intensive culture system might help to express their maximum growth although there were differences in the salinity in both extensive (14.9 ± 6.93 ppt) and intensive (21.6 ± 7.66 ppt) systems.

7.6 Conclusions

As the number of tilapia breeding programme increases, and the introduction of new and genetically improved fishes expands, it is important to develop a quick and cost effective method(s) to estimate the genetic purity of tilapia species. A set of species-specific SNP markers was identified, which were found to be diagnostic for *O. mossambicus, O. niloticus*, and *O. aureus*, the commercially most important aquaculture species. These SNP markers were efficient to identify the genomic composition of GIFT (which was developed from several mixed populations) and the Molobicus hybrid (developed from crosses between GIFT and feral *O. mossambicus*). This study also estimated the genomic constitution of the Molobicus strain and concluded that the genome contribution moved towards *O. niloticus* compared to *O. mossambicus* after seven generations of selective breeding in which higher harvest weight was the selection criteria. This pilot study with 10 species-diagnostic SNP markers adds a first case study demonstrating the potential of this approach for the genetic management of tilapia species through identifying the pure and hybrid species.

Most of the genomic regions (out of 10 markers assayed) showed a move towards the *O. niloticus* in the Molobicus generation 07 might be due to selection for higher harvest weight but some moved in the opposite direction. It doesn't seem to be so clear due to complications in GIFT composition, but some regions might be expected to link to salinity tolerance and to move towards *O. mossambicus*. More species diagnostic markers across the whole genome need to be identified using RADseq/ddRADseq, or whole genome sequencing of the *O. mossambicus* (once *O. mossambicus* genome assembly is available, currently underway in Singapore) followed by aligning the sequences to *O. niloticus* (published genome assembly Orenil1.1; Brawand *et al*. 2014) to identify the more species

diagnostic markers which will be helpful to the genetic management and genetic improvement of the widely cultured tilapia species.

Acknowledgements

The authors acknowledge the Commonwealth Scholarship Commission (UK) for their financial assistance (CSC reference BDCA-2013-5). The work was also partly supported by the MASTS pooling initiative (The Marine Alliance for Science and Technology for Scotland) funded by the Scottish Funding Council (grant reference HR09011) and contributing institutions.

Chapter 8

Discussion, conclusions and future directions

This research project was conducted to elaborate the existing knowledge of the genetic bases underlying sex determination and to further improve the genetic management of Nile tilapia stocks/strains using genomic approaches based on high-throughput sequencing technologies. One of the major objectives was to develop a methodology to allow screening of populations and individual fish for their genetic sex, particularly in genetically improved farmed tilapia (GIFT), with the analysis starting at the population level (multiple families) using bulk segregant analysis (BSA), given the emerging evidence for variation in sex determination system between and within different populations of Nile tilapia. Another important objective was to apply more efficient marker-assisted selection, using tightly sex-linked markers, in Nile tilapia for hormonal feminisation, an essential but problematic step in the production of genetically male tilapia. A set of species-diagnostic SNP markers for tilapia was also verified and tested in case studies, along with a minimally invasive DNA sampling technique, both of which will be helpful in the genetic management of tilapia species and stocks.

This last and concluding chapter will discuss the major findings of this PhD research chronologically and will then suggest future research directions.

8.1 Nile tilapia sex determination (Stirling population)

It has been established that the Nile tilapia sex is mostly determined by the major sex-determining loci with some minor genetic/environmental influences. Two different major (XX/XY) sex-determining loci (LG1 and LG23) have been identified in Nile tilapia. LG1 has been found to be the major sex-determining locus in the Stirling stock with some thermosensitivity modification from LG20; no previous evidence of any LG23 influence had been detected before the present study. The present study (Chapter 5) revealed a rare

polymorphism in LG23 that was associated with being phenotypic male in the Stirling stock. One informative Stirling family for the LG23 locus (non-informative for LG1 and LG20) was identified, and a strong sex-LG23 association was found in the progeny. Two families informative for both LG1 and LG23 markers were produced, but the segregation pattern is yet to be determined (not analysed by the time of submission of this thesis).

Recent studies show the complexity of the sex determination in different Nile tilapia stocks, suggesting additional difficulties in developing all-male tilapia production systems by using a marker-assisted approach, because tightly sex-linked markers in one strain or stock may not be sex-linked in another population. Tightly sex-linked markers in LG1 (both SNP and microsatellite; Palaiokostas *et al*. 2013a) could be used to distinguish between most phenotypically male and female fish in the Stirling Nile tilapia population, although the results from the present study suggest that fish should also be screened using LG23 marker(s). These markers were used in the study on hormonal feminisation using estrogen hormones and high temperature in families with different sexual genotypes (Chapter 6). The combined treatment of estrogen hormone and high temperature during the sex differentiation period in Nile tilapia induced higher feminisation than hormone alone in all sex groups, although estrogen hormone and high temperature were thought to have an antagonistic action in the sex differentiation pathway in fish species including Nile tilapia (Kobayashi *et al*. 2003, 2008; Wessels and Hörstgen-Schwar 2007). DES was found to be more efficient than EE2 hormone but the survival rate was higher in EE2 treated batches compared to DES. The next step would be crossing these XY neo-females with YY males and treating the progeny with different types/doses of estrogen hormones (with elevated temperature) to optimize the protocol for producing YY neo-females, which in turn should produce all YY male progeny when crossed with YY males. Different types of YY and XY fish can be produced in Stirling stock, as the LG1 is the major sex-determining locus and

rare LG23 influence is also present. Now there is a question what genotype - "LG1 YY's, LG23 YY's or double (LG1 + LG23) YY's" would give us better results on the way to produce all male production. So the next step would be to develop (including hormonal feminisation) and test (sex-ratios, performance) both types using LG1 and LG23 markers.

8.2 GIFT tilapia sex determination

Genetically improved farmed tilapia (GIFT) is a synthetic line of Nile tilapia, developed through selective breeding by the ICLARM/WorldFish Center since 1988 from a wide variety of wild (African) and farmed (Asian) populations of Nile tilapia. Different major sex-determining loci have been identified in different populations of Nile tilapia, and the domesticated populations used for GIFT development have been reported to be introgressed with *Oreochromis mossambicus*. Given the mixing of different populations of tilapia (probably possessing sex QTL in different chromosomes and with evidence of gene introgression from other species) to develop a new GIFT strain, sex determination in GIFT tilapia was unpredictable.

Given the evidence of complexity of sex determination in Nile tilapia and the fact that previous studies had been conducted with only a few families, a major objective was to analyse sex determination in GIFT at the population level (Chapter 4). The original plan was to use 30 GIFT families, however this number of families was not available with adequate numbers of progeny. For these reasons, 19 GIFT families and 50 broodstock were analysed. Constructing a double-digest restriction-site associated DNA (ddRAD) library using individual samples from 19 families would be expensive, time consuming and very tedious. With a view to reducing the large number of samples and processing costs and time, a bulk segregant analysis (BSA) approach to ddRADseq was taken. To further reduce

the cost and labour of individual DNA extraction, we extended the BSA approach by pooling of equal amount of tissue samples for each progeny pool followed by DNA extraction for ddRADseq.

This was the first genomic analysis of sex determination in GIFT, and uncovered the underlying mechanism across the population. Surprisingly SNP markers with high linkage probability to the phenotypic sex clustered in only a single chromosome (LG23) in the majority of the GIFT families in the BSA-ddRADseq analysis. No other QTL linked to phenotypic sex was detected from the analysis. More importantly the same association persisted across the families and population in the later individual analyses (6 families - 230 progenies and 50 broodstock) using 14 LG23 DNA-based sex-linked markers (5 SNPs from BSA-ddRAD analysis; 2 microsatellites; 2 insertions, 3 deletions and 2 SNPs in the variants of the *Amh* gene). With very few exceptions, females and males were homozygous or heterozygous respectively for the SNPs (from BSA-ddRAD analysis) with highest association with the phenotypic sex. Based on the microsatellite marker (LG23) analysis, two alleles 267 (for *UNH898*) and 274 (for *ARO172*) were always associated with the male phenotype.

The availability of a high quality Nile tilapia genome assembly assisted in the identification of the candidate sex-determining gene in some Nile tilapia stocks. Eshel *et al*. (2014) first identified a duplicated variant of the *Amh* gene on the Y-chromosome, which was found to be associated with the male sex in an "Israeli" strain of Nile tilapia. Later on Li *et al.* (2015) also identified the same variant along with another tandemly located variant of the *Amh* gene which was associated with the male sex in the Japanese population of Nile tilapia, and was proposed to be the male sex-determining gene. Neither Eshel *et al.* (2014) or Li *et al.* (2015) have demonstrated this at the population level. The present study analysed the Y-

linked variants of the *Amh* gene in GIFT at the population level and these were found to be linked to male sex determination. The *AmhΔy* variant in GIFT was exactly the same as *AmhΔy* observed by Li *et al.* (2015) in the Japanese strain, as far as could be determined, whereas differences were observed for *Amhy*; a previously unknown three base pair insertion in Exon 0 and no polymorphism in exon II (which was thought to be the key regulator for male sex determination in Japanese population), which suggest that *Amhy* in GIFT might be another variant of the *Amh* gene, and a candidate gene for male sex determination in this strain. No attempt was made in this study to determine if either of these *Amh* variants were the actual sex-determining gene(s) in GIFT.

The sex-linked markers identified would allow marker-assisted selection, for the first time, in GIFT removing much of the need for the long process of progeny testing to confirm the genotypes. The next step would be using those sex-linked markers to control the sex-ratio in the culture system by producing all-male GIFT tilapia which could prevent the problems caused by reproduction before harvest in Nile tilapia. Sex in Nile tilapia may also be influenced by temperature, but elevated temperature has not been found to have any effect on sex-ratio in GIFT (unpublished, personal communication John Benzie). The present study has demonstrated the association of the *Amh* gene variants in LG23 with male sex determination in GIFT. It would be interesting to analyse the expression level of *Amh*, the candidate gene, or other genes involved in the sex differentiation pathway in both male and female GIFT tilapia.

Genotyping of tens to hundreds of thousands of SNP markers on a genome-wide scale in a single chip-based assay has emerged as an attractive tool in the field of genetics and genomics over the last few years (Steermers and Gunderson 2007). About 6.5 K to 15 K SNP-chips have been designed to study population genetic structure and QTL mapping for

Atlantic salmon (Dominik *et al*. 2010; Gutierrez *et al*. 2012). In the case of terrestrial agricultural animals, a range of SNP-chips are available, e.g. for pig and chicken (Ramos *et al*. 2009; Groenen *et al*. 2011). However, only four SNPs (out of 9,108) were found to be informative/common in all 19 GIFT families (**Table 8.1**), which is not adequate or at least not ideal to design a SNP-chip. This low number could perhaps be attributed to the multiple strain origin of the GIFT population and the level of polymorphism among the families.

Table 8.1 Number of common SNPs (total unique SNPs) in 19 GIFT families.

	Number of unique SNPs	**Number of families**
	2,467	1
	1,360	2
	1,012	3
	687	4
	560	5
	471	6
	387	7
	338	8
	317	9
	320	10
	245	11
	239	12
	218	13
	189	14
	145	15
	86	16
	47	17
	16	18
	4	19
Total	9,108	19

8.2.1 Pre-extraction pooling approach to ddRADseq

BSA is a rapid gene mapping technique, based on two phenotypically contrasting pooled samples that differ for a trait of interest, and with the rapid improvement of the NGS techniques, BSA has recently been combined with different NGS approaches (Trick *et al*. 2012; Livaja *et al*. 2013). Pooling of equal amounts of DNA for BSA is a widely used method. There are some constraints that have been reported for BSA-mediated NGS, especially restriction-based studies such as RADseq (Gautier *et al*. 2013a, b; Schlötterer *et al*. 2014). In our study, an equal amount of tissue sample from individuals of the same phenotypic sex was pooled before extraction and used for ddRADseq, which might be challenging because all individuals might not be represented equally.

As tissue samples were pooled, two Stirling families were produced and their genomic sex was confirmed using tightly sex-linked markers in LG1 and LG20 before preparing samples for BSA-ddRADseq and using these as positive controls (Chapter 4). BSA-ddRADseq analysis also identified the same genomic position (LG1) showing very strong association with phenotypic sex in both Stirling families, which means that the pre-extraction pooling of tissue samples to ddRADseq was able to identify the genomic region linked to the trait of interest (sex in this case). Approximately 1,500 – 3,000 bi-allelic polymorphic loci per family were identified from the BSA-ddRADseq dataset. Palaiokostas *et al*. (2015) retrieved approximately 1,200 bi-allelic polymorphic loci from the individual analysis in Nile tilapia using ddRADseq. GIFT families were also analysed using the same approach and only one locus (LG23) was found to be associated with phenotypic sex. Some GIFT families showed weaker association and a few families did not show any specific association. These with weaker or no association could be because of errors in phenotypic sex identification (detailed explanation in Chapter 4), unequal representation of each

individual in ddRADseq pools or minor genetic/environmental factors affecting sex in GIFT.

The BSA-ddRADseq approach allowed mapping of a single locus affecting a trait of interest. In the case of a trait which is controlled by two loci on different chromosomes (as seems possible in some stocks of Nile tilapia) it remains to be seen whether BSA-ddRADseq would be able to detect the association with both loci. However, it does seem likely that, if sex is controlled by two QTLs in chromosome 1 and 23, and BSA pools are constructed according to the phenotypic sex for ddRADseq, it would be possible to detect the associations of both loci with phenotypic sex (**Figure 8.1**).

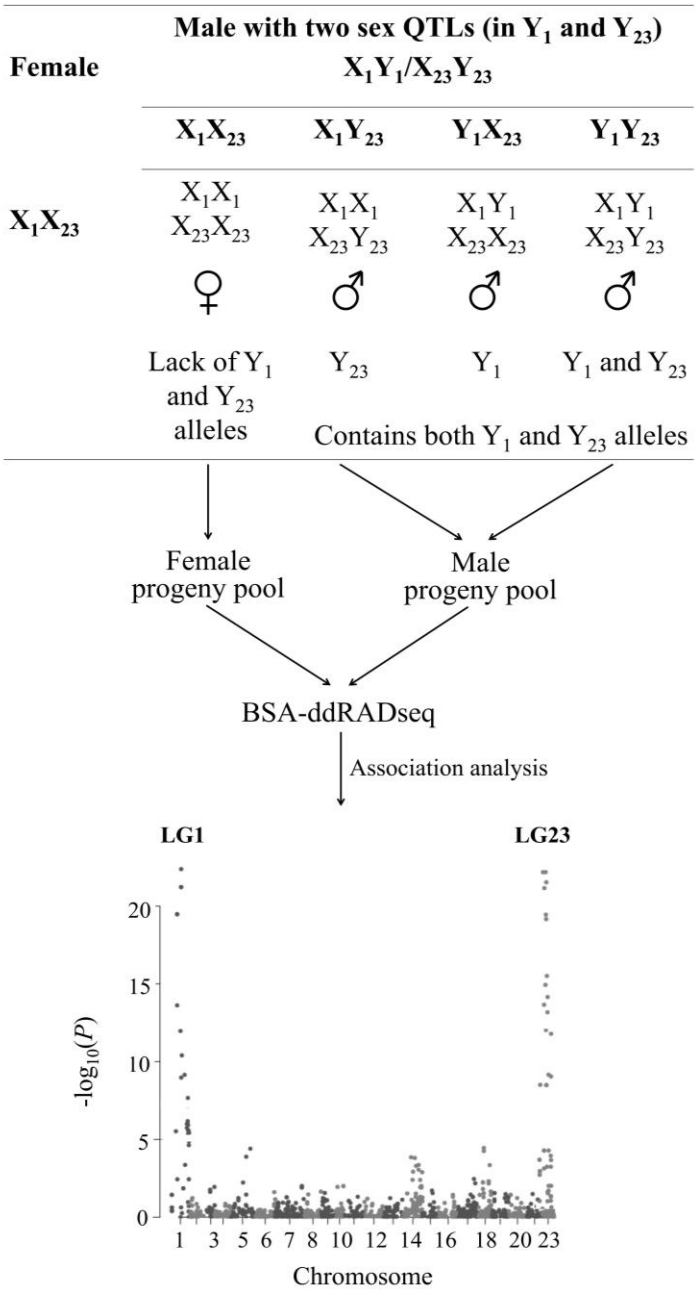


Figure 8.1 Hypothetical outcome if sex is determined by two QTLs (LG1 and LG23). It appears possible to find association with both loci using the BSA-ddRADseq approach.

This cost-effective approach will provide useful information for genomic studies and will allow sequencing of lots of individuals from a single population or multiple populations in a

single sequencing run to identify a single locus (present study) or two loci (yet to be investigated) influencing commercially important traits.

8.3 Mystery of tilapia sex determination

Evolution of sex determination has played a fundamental role in the diversification of African Cichlid fishes. Complex sex-determining systems have been observed in haplochromine cichlids in Lake Malawi, including a major female heterogametic locus on LG5 (WZ), two male heterogametic loci on LG7 (XY) and two interacting loci on LG3 (WZ) and LG20 (XY) (Roberts *et al*. 2009; Ser *et al*. 2010; Parnell and Streelman 2013). Tilapia also exhibit both male (XY) and female (WZ) heterogametic systems, and the major sex-determining loci have been mapped to LG1 (XY), LG3 (WZ) and LG23 (XY). Multiple loci sometimes control the sex determination in a single species. For instance, interaction between the LG3 ZW system and LG1 XY system appears to determine sex in *O. aureus*, although based on only one family (Lee *et al*. 2004).

Nile tilapia has a male heterogametic sex-determining system, and inter and intra-population variations have been observed; some possess a major sex-determining locus in LG1 while in others it is LG23. In the University of Stirling population of Nile tilapia derived from Lake Manzala in Egypt, LG1 was found to be the major sex-determining locus using both microsatellite markers and RADseq-based techniques (Lee *et al*. 2003; Palaiokostas *et al*. 2013a). More recently a different locus in LG20 has been found to be linked to male sex determination in some thermosensitive families in the same stock (Palaiokostas *et al*. 2015). Our study demonstrated a rare Y-linked LG23 locus also influences sex in the Stirling population (Chapter 5).

In an Israeli population of Nile tilapia (Agricultural Research Organization), derived from Stirling via Swansea, a major sex-determining locus was identified in LG23 (Eshel *et al*. 2011, 2012). Eshel *et al*. (2014) identified a duplicated copy of *Amh* gene in LG23 associated with male sex determination in this population. Very recently two duplicated copies of Y-linked *Amh* gene (*AmhΔy*, *Amhy*) have been identified in the Japanese strain (National Institute for Basic Biology) of Nile tilapia originally from Egypt, Africa. In GIFT, interestingly only a single locus (LG23) appears to determine sex across the population despite the mixed origins of the strain (Chapter 4). GIFT had exactly the same copy of *AmhΔy* variant of *Amh* gene but differences were observed in the case of *Amhy*.

In a German population of Nile tilapia (University of Göttingen) derived from Stirling, polymorphisms have been identified in LG1, 3 and 23 which were linked to male sex determination in some of the high temperature treated families, from a line selected for greater response to temperature (Lühmann *et al*. 2012). Wessels *et al*. (2014) also identified an allelic variant in *Amh* gene, which was associated with temperature-induced sex-reversal in the same population of Nile tilapia.

It is clear now that two major XX/XY loci (in LG1 and LG23) determine sex in different populations of Nile tilapia and LG23 looks like the more common locus. The Stirling Nile tilapia population was brought from Lake Manzala Egypt in 1979 to the University of Stirling, which in turn was distributed to different countries (Israel via Swansea and Germany) from this small founder population. LG1 is the major sex-determining locus in Stirling population (small effect from LG23) while LG23 (*Amh* gene) is the major sex-determining locus in the Israel and in the Japanese population derived from Egypt. Due to founder and bottleneck effect, the LG1 XX/XY locus could have been lost in the Israel population and LG23 remained as the pre-dominant one. Loss of the sex locus has been

observed in domesticated populations of Zebrafish (model species), whereas in wild Zebrafish a major (WZ/ZZ) sex-determining locus has been found in chromosome 4, suggesting that the modification/loss in the sex-determining locus happened during the domestication process (Wilson *et al*. 2014). Another explanation could be that the same gene is determining sex in different populations of Nile tilapia (major sex-determining gene/genes are yet to be explored thoroughly). It has been recorded in salmonids that a single master sex-determining gene *Sdy* (associated with transposable elements) was found to be conserved across the salmonids and jumped into different chromosomes during the evolution of salmonids. In a farmed Tasmanian Atlantic salmon population, three different sex-determining loci (Ssa02, Ssa03, Ssa06) have been identified in three different chromosomes in the same population (Eisbrenner *et al*. 2014). The sex-determining gene *Sdy* (conserved in salmonids) was also evident in the same Tasmanian salmon population, which suggest that the position of this gene might be in three different chromosomes. Given that the Nile tilapia do have a candidate gene in LG23 (*Amh* variant[s]) and there is no sign of this in LG1. Therefore we would have to postulate that actually *Amh* variant(s) are not the main male sex-determining genes in Nile tilapia (some doubt from comparison of GIFT *Amhy* sequence data and that of Li *et al*. 2015); but another tightly linked gene or regulatory factor might be present in both LG23 and LG1.

It would be very interesting to study sex determination in wild Nile tilapia populations from which the domesticated ones originated (if they still exist - purity would be another concern) to compare with the domesticated populations, to see whether they demonstrate the same sex-determining mechanisms.

8.4 Genetic management of tilapia species

Genetic management of populations in the wild or in captivity is a crucial process to ensure that the populations are genetically viable, healthy and sustainable across the generations. In the case of endangered species, tissue biopsy could have a negative impact on the welfare and survival of sampled individuals. In countries where a licence is required to work with animals, the development of non-invasive or minimally invasive DNA sampling, which does not require licencing, would simplify the whole management process.

A minimally invasive DNA sampling method using fish mucus was developed, and tested for standard genotyping and ddRADseq (Chapter 3). Initially mucus DNA was assessed for microsatellite and SNP marker analysis and was found to be suitable for both analyses, although significant amounts of bacterial DNA could be amplified from the mucus DNA. Whether this bacterial DNA has any negative impact on the output from the ddRADseq was also investigated. Mucus-derived DNA generated similar quality and quantity of data from ddRADseq as DNA derived from other sources (invasive tissue biopsy - fin or muscle). No bacterial sequences were observed in the mucus-derived DNA sequence data. Such a minimally invasive technique will increase the applicability of NGS in a wide range of molecular science especially for conservation genetics. Even though mucus-derived DNA was successfully used for standard genotyping and ddRADseq, it could be of much interest to make a comparison between mucus and fin-derived DNA for parentage assignment, QTL mapping or other types of genetic analysis.

With the widespread movements of tilapia species and the difficulties of identifying some of these on the basis of morphology, which leads to identify the specific problem for properly managing breeding programme. Several fish breeding programmes on Nile tilapia have been implemented and GIFT has been widely used as a base population for majority of

those breeding programmes, it is important to know the genetic purity of the base population before initiating any breeding programme. A set of species-specific diagnostic markers for different tilapia species has been designed from a ddRADseq analysis of several species (Syaifudin 2015). Among them 10 SNPs (four with an allele diagnostic for *O. niloticus*, four for *O. mossambicus* and two for *O. aureus*) were selected and first validated on reference material from pure species (Chapter 7). This confirmed the species-specificity of the 10 SNP markers (one *O. niloticus* marker with a rare allele).

It was found that the major proportion of the GIFT genome was from Nile tilapia with a small proportion from *O. mossambicus* based on the 10 species-diagnostic SNP markers. Previous studies reported that some of the base populations used for GIFT development were introgressed with *O. mossambicus* (Taniguchi *et al*. 1985; Macaranas *et al*. 1986; Eknath *et al*. 1991). It would be interesting future work to analyse the base population used for GIFT development (if samples are available) using species-diagnostic SNP markers, or the individual Asian Nile tilapia base stocks that are thought to have had some introgression with *O. mossambicus.*

"Molobicus" is a hybrid tilapia developed through selective breeding programme (based on crosses between GIFT and feral *O. mossambicus*), aimed to develop a tilapia strain with high growth performance in saline environment. Samples from *O. mossambicus* (parent for Molobicus strain), F1 hybrid and generation 7 were available. Feral *O. mossambicus* was found to be nearly pure *O. mossambicus* and Molobicus F1 hybrid was found to be hybrid with some deviations (nearly all individuals were heterozygote for the 10 markers analysed). When compared with the samples from generation 7, significant deviation in the species allele frequency ratio from the expected one was observed which suggest that the after seven generations of selective breeding based on the highest harvest weight, the

Molobicus strain has significantly increased proportion of Nile tilapia genome at the expense of *O. mossambicus*. It would add more clear information if we could have analysed the actual GIFT parent (from 7th generation) used for the Molobicus strain. Only 10 SNPs were used for this study, which do not represent the whole genome. NGS techniques offer great potential to generate massive number of SNPs and it would be interesting if a large number of diagnostic SNP markers across the whole genome could be used, or from resequencing and compare to the two genome wide sequencing (WGS) assemblies (once the *O. mossambicus* genome available) from the two parent species.

8.5 Conclusions

Nile tilapia sex determination varies between and within populations and the actual sex-determination in a given population needs to be explored. This PhD research discovered that a single sex-determining locus (potentially a candidate gene *Amh*) controls sex in GIFT tilapia, which is surprising given the multiple origins of this species. The sex-linked markers identified will be helpful to apply marker-assisted selection in GIFT to produce all-male tilapia to prevent reproduction in culture systems such as ponds. This study also identified the existence of an LG23 XX/XY sex-determining locus, with a different Y allele to GIFT, in the Stirling Nile tilapia population (as found to be the predominant locus in an Israeli population, derived from Stirling through Swansea). This, in addition to the LG1 XX/XY locus, helps to explain the previous problems in the production of YY males in the Stirling population, and will help towards more efficient MAS in this population. Feminization in Nile tilapia is harder to induce than masculinization. The combined treatment with estrogen hormones and high temperature proved to be efficient to increase the feminization rate compared to hormone alone, which will be useful to produce XY and YY neo-females, important in generating all-male tilapia. Several tilapia species-diagnostic

SNP markers were tested in putative pure species and fish from two breeding programmes with input from at least two species (GIFT and Molobicus), to assess how the genomes of these fish had been reshaped through several generations of selective breeding. Such markers can be applied in a variety of aquaculture and wild fish populations to address hybridization and introgression issues, which are widespread in tilapias. This research also developed two methods which should have a range of applications: minimally invasive mucus DNA sampling for standard genotyping and ddRAD analysis, and pre-extraction pooling of tissue samples to ddRADseq for bulk segregant analysis. Overall this research provides novel information on the complex sex-determining system(s) in Nile tilapia and demonstrates the power of new SNP markers, which have potential to contribute to improved genetic management of tilapia species in world aquaculture production.

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.

Adépo-Gourène B., Gourène G., Agnèse J.F. 2006. Genetic identification of hybrids between two autochthonous tilapia species, *Tilapia zillii* and *Tilapia guineensis*, in the man-made lake Ayame. Aquatic Living Resources, 19(3):239-245.

Admassu D., Casselman J.M. 2000. Otolith age determination for adult tilapia, *Oreochromis niloticus* L. from Lake Awassa (Ethiopian Rift Valley) by interpreting biannuli and differentiating biannual recruitment. Hydrobiologia, 418(1):15-24.

Agnése J.F., Adépo-Gourène B., Owino J., Pouyaud L., Aman R. 1999. Genetic characterization of a pure relict population of *Oreochromis esculentus*, an endangered tilapia. Journal of Fish Biology, 54(5):1119-1123.

Akbari M., Hansen M.D., Halgunset J., Skorpen F., Krokan H.E. 2005. Low copy number DNA template can render polymerase chain reaction error prone in a sequence-dependent manner. The Journal of Molecular Diagnostics, 7(1):36-39.

Al-Amoudi M.M. 1987. Acclimation of commercially cultured *Oreochromis* species to sea water - an experimental study. Aquaculture, 65:333-342.

Al-Attar L., Noël K., Dutertre M., Belville C., Forest M.G., Burgoyne P.S., Josso N., Rey R. 1997. Hormonal and cellular regulation of Sertoli cell anti-Müllerian hormone production in the postnatal mouse. Journal of Clinical Investigation, 100(6):1335-1343.

Aljanabi S.M., Martinez I. 1997. Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. Nucleic Acids Research, 25(22):4692-4693.

Altschul S.F., Gish W., Miller W., Myers E.W., Lipman D.J. 1990. Basic local alignment search tool. Journal of Molecular Biology, 215:403-410.

Amberg J.J., Goforth R., Stefanavage T., Sepúlveda M.S. 2010. Sexually dimorphic gene expression in the gonad and liver of shovelnose sturgeon (*Scaphirhynchus platorynchus*). Fish Physiology and Biochemistry, 36(4):923-932.

Amores A., Catchen J., 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(4):799-808.

Amos C.I., Frazier M.L., Wang W. 2000. DNA pooling in mutation detection with reference to sequence analysis. The American Journal of Human Genetics, 66(5):1689-1692.

Angelopoulou R., Lavranos G., Manolakou P. 2012. Sex determination strategies in 2012: towards a common regulatory model? Reproductive Biology and Endocrinology, 10(1):13.

Angienda P.O., Lee H.J., Elmer K.R., Abila R., Waindi E.N., Meyer A. 2011. Genetic structure and gene flow in an endangered native tilapia fish (*Oreochromis esculentus*)

compared to invasive Nile tilapia (*Oreochromis niloticus*) in Yala swamp, East Africa. Conservation Genetics, 12(1):243-255.

Arakawa Y., Nishida-Umehara C., Matsuda Y., Sutou S., Suzuki H. 2002. X-chromosomal localization of mammalian Y-linked genes in two XO species of the Ryukyu spiny rat. Cytogenetic and Genome Research, 99:303-309.

Argue B.J., Arce S.M., Lotz J.M., Moss S.M. 2002. Selective breeding of pacific white shrimp (*Litoåenaeus vannamei*) for growth and disease resistance to Taura Syndrome. Aquaculture, 204:447-460.

Arnheim N., Strange C., Erlich H. 1985. Use of pooled DNA samples to detect linkage disequilibrium of polymorphic restriction fragments and human disease: studies of HLA class II loci. Proceedings of the National Academy of Sciences of the United States of America, 82(20):6970-6974.

Atz J. W. 1964. Intersexuality in fishes. Academic Press.

Avise J.C. 1994. Molecular markers: natural history and evolution, Chapman and Hall, New York.

Aylmer F.R. 1930. The genetical theory of natural selection. Oxford University Press.

Azaza M.S., Dhraïef M.N., Kraïem M.M. 2008. Effects of water temperature on growth and sex ratio of juvenile Nile tilapia *Oreochromis niloticus* (Linnaeus) reared in geothermal waters in southern Tunisia. Journal of Thermal Biology, 33(2):98-105.

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(10):e3376.

Baker B.S., Wolfner M.F. 1988. A molecular analysis of doublesex, a bifunctional gene that controls both male and female sexual differentiation in *Drosophila melanogaster*. Genes & Development, 2(4):477-489.

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

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

Baroiller J.F., D'Cotta H. 2001. Environment and sex determination in farmed fish. Comparative Biochemistry and Physiology Part C: Toxicology & Pharmacology, 130(4):399-409.

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

Baroiller J.F., Guiguen Y., Fostier A. 1999. Endocrine and environmental aspects of sex differentiation in fish. Cellular and Molecular Life Sciences, 55(6):910-931.

Baron D., Cocquet J., Xia X., Fellous M., Guiguen Y., Veitia R.A. 2004. An evolutionary and functional analysis of *Foxl2* in rainbow trout gonad differentiation. Journal of Molecular Endocrinology, 33(3):705-715.

Bastide H., Betancourt A., Nolte V., Tobler R., Stöbe P., Futschik A., Schlötterer C. 2013. A genome-wide, fine-scale map of natural pigmentation variation in *Drosophila melanogaster*. Plos Genetics, 9(6):e1003534.

Baxter S.W., Davey J.W., Johnston J.S., Shelton A.M., Heckel D.G., Jiggins C.D., Blaxter M.L. 2011. Linkage mapping and comparative genomics using next-generation RAD sequencing of a non-model organism. Plos One, 6(4):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.

Bentsen H.B., Eknath A.E., Palada-de Vera M.S., Danting J.C., Bolivar H.L., Reyes R.A., Dionisio E.E., Longalong F.M., Circa A.V., Tayamen M.M., Gjerde B. 1998. Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of *Oreochromis niloticus*. Aquaculture, 160:145-173.

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, 272 (Suppl. 1):S3-S16.

Bláquez M., Zanuy S., Carillo M., Piferrer F. 1998. Effects of rearing temperature on sex differentiation in the European sea bass (*Dicentrarchus labrax* L.). Journal of Experimental Zoology, 281(3):207-216.

Blázquez M., Somoza G.M. 2010. Fish with thermolabile sex determination (TSD) as models to study brain sex differentiation. General and Comparative Endocrinology, 166(3):470-477.

Blears M.J., De Grandis S.A., Lee H., Trevors J.T. 1998. Amplified fragment length polymorphism (AFLP): a review of the procedure and its applications. Journal of Industrial Microbiology & Biotechnology, 21(3):99-114.

Böhne A., Heule C., Boileau N., Salzburger W. 2013. Expression and sequence evolution of aromatase *Cyp19a1* and other sexual development genes in East African cichlid fishes. Molecular Biology and Evolution, 30(10):2268-2285.

Bondary K. 1983. Response to bidirectional selection for body weight in channel catfish. Aquaculture, 33:73-81.

Botstein D., White R.L., Skolnick M., Davis R.W. 1980. Construction of a genetic linkage map in man using restriction fragment length polymorphisms. American Journal of Human Genetics, 32(3):314-331.

Boutin-Ganache I., Raposo M., Raymond M., Deschepper C.F. 2001. M13-tailed primers improve the readability and usability of microsatellite analyses performed with two different allele-sizing methods. BioTechniques, 31(1):24-28.

Bouza C., Hermida M., Pardo B.G., Fernández C., Fortes G.G., Castro J., Sánchez L., Presa P., Pérez M., Sanjuán A., de Carlos A., Álvarez-Dios J.A., Ezcurra S., Cal R.M., Piferrer F., Martínez P. 2007. A microsatellite genetic map of the turbot (*Scophthalmus maximus*). Genetics, 177(4):2457-2467.

Brawand D., Wagner C.E., Li Y.I., Malinsky M., Keller I., Fan S. *et al.* 2014. The genomic substrate for adaptive radiation in African cichlid fish. Nature, 513:375-381.

Brown J.K., Taggart J.B., Bekaert M., Wehner S., Palaiokostas C., Setiawan A.N., Symonds J.E., Penman D.J. 2016. Mapping the sex determination locus in the hāpuku (*Polyprion oxygeneios*) using ddRAD sequencing. BMC Genomics, 17:448.

Buckland-Nicks J.A., Gillis M., Reimchan T.E. 2012. Neural network detected in a presumed vestigial trait: ultrastructure of the salmonid adipose fin. Proceedings of the Royal Society of London B: Biological Sciences, 279(1728):553-563.

Budd A.M., Banh Q.Q., Domingos J.A., Jerry D.R. 2015. Sex control in fish: approaches, challenges and opportunities for aquaculture. Journal of Marine Science and Engineering, 3(2):329-355.

Cambray J., Swartz E. 2007. *Oreochromis mossambicus*. The IUCN Red List of Threatened Species. www.iucnredlist.org.

Campanella J.J., Smalley J.V. 2006. A minimally invasive method of piscine tissue collection and an analysis of long-term field-storage conditions for samples. BMC Genetics, 7:32.

Campos-Ramos R., Harvey S.C., Masabanda J.S., Carrasco L.A.P., Griffin D.K., McAndrew B.J., Bromage N.R., Penman D.J. 2001. Identification of putative sex chromosomes in the blue tilapia, *Oreochromis aureus*, through synaptonemal complex and FISH analysis. Genetica, 111:143-153.

Canonico G.C., Arthington A., McCrary J.K., Thieme M.L. 2005. The effects of introduced tilapias on native biodiversity. Aquatic Conservation: Marine and Freshwater Ecosystems, 15(5):463-483.

Carmi R., Rokhlina T., Kwitek-Black A.E., Elbedour K., Nishimura D., Stone E.M., Sheffield V.C. 1995. Use of a DNA pooling strategy to identify a human obesity syndrome locus on chromosome 15. Human Molecular Genetics, 4(1):9-13.

Catchen J., Hohenlohe P.A., Bassham S., Amores A., Cresko W.A. 2013. Stacks: an analysis tool set for population genomics. Molecular Ecology, 22:3124-3140.

Cate R.L., Mattaliano R.J., Hession C., Tizard R., Farber N.M., Cheung A., Ninfa E.G., Frey A.Z., Gash D.J., Chow E.P., Fisher R.A., Bertonis J.M., Torres G., Wallner B.P., Ramachandran K.L., Ragin R.C., Manganaro T.F., MacLaughlin D.T., Donahoe P.K. 1986. Isolation of the bovine and human genes for Müllerian inhibiting substance and expression of the human gene in animal cells. Cell, 45(5):685-698.

Chang X., Kobayashi T., Senthilkumaran B., Kobayashi-Kajura H., Sudhakumari C.C., Nagahama Y. 2005. Two types of aromatase with different encoding genes, tissue distribution and developmental expression in Nile tilapia (*Oreochromis niloticus*). General and Comparative Endocrinology, 141(2):101-115.

Chen S.L., Li J., Deng S.P., Tian Y.S., Wang Q.Y., Zhuang Z.M., Sha Z.X., Xu J.Y. 2007. Isolation of female-specific AFLP markers and molecular identification of genetic sex in half-smooth tongue sole (*Cynoglossus semilaevis*). Marine Biotechnology, 9(2):273-280.

Chiang E.F.L., Pai C.I., Wyatt M., Yan Y.L., Postlethwait J., Chung B.C. 2001. Two *Sox9* genes on duplicated Zebrafish chromosomes: expression of similar transcription activators in distinct sites. Developmental Biology, 231(1):149-163.

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 microsatellite linkage map of the European sea bass *Dicentrarchus labrax* L. Genetics, 170(4):1821-1826.

Chue J., Smith C.A. 2011. Sex determination and sexual differentiation in the avian model. The FEBS Journal, 278(7):1027-1034.

Clarke A.C., Prost S., Stanton J.A.L., White W.T.J., Kaplan M.E., Matisoo-Smith E.A. 2014. From cheek swabs to consensus sequences: an A to Z protocol for high throughput DNA sequencing of complete human mitochondrial genomes. BMC Genomics, 15:68.

Cline T.W. 1993. The Drosophila sex determination signal: how do flies count to two? Trends in Genetics, 9(11):385-390.

Cnaani A., Hallerman E.M., Ron M., Weller J.I., Indelman M., Kashi Y., Gall G.A.E., Hulata G. 2003. Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F2 tilapia hybrid. Aquaculture, 223:117-128.

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. 2008. Genetics of sex determination in tilapiine species. Sexual Development, 2(1):43-54.

Cnaani A., Zilberman N., Tinman S., Hulata G., Ron M. 2004. Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. Molecular Genetics and Genomics, 272(2):162-172.

Conover D.O., Heins S.W. 1987. Adaptive variation in environmental and genetic sex determination in a fish. Nature, 326(6112):496-498.

Conover D.O., Kynard B.E. 1981. Environmental sex determination: interaction of temperature and genotype in a fish. Science, 213(4507):577-579.

Costa-Pierce B.A. 2003. Rapid evolution of an established feral tilapia (*Oreochromis* spp.): the need to incorporate invasion science into regulatory structures. In: Marine Bioinvasions: Patterns, Processes and Perspectives Springer Netherlands (pp. 71-84).

Craig J.K., Foote C.J., Wood C.C. 1996. Evidence for temperature-dependent sex determination in sockeye salmon (*Oncorhynchus nerka*). Canadian Journal of Fisheries and Aquatic Sciences, 53(1):141-147.

Cresko W.A., Yan Y.L., Baltrus D.A., Amores A., Singer A., Rodríguez-Marí A., Postlethwait J.H. 2003. Genome duplication, subfunction partitioning, and lineage divergence: *Sox9* in stickleback and Zebrafish. Developmental Dynamics, 228(3):480-489.

Crews D., Bergeron J.M., Bull J.J., Flores D., Tousignant A., Skipper J.K., Wibbels T. 1994. Temperature-dependent sex determination in reptiles: Proximate mechanisms, ultimate outcomes, and practical applications. Developmental Genetics, 15(3):297-312.

Cummings S.A., Thorgaard G.H. 1994. Extraction of DNA from fish blood and sperm. BioTechniques, 17:426-428.

Cutting A., Chue J., Smith C.A. 2013. Just how conserved is vertebrate sex determination? Developmental Dynamics, 242(4):380-387.

D'Amato M.E., Esterhuyse M.M., van der Waal B.C.W., Brink D., Volckaert F.A.M. 2007. Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping. Conservation Genetics, 8(2):475-488.

D'Cotta H., Pepey E., Pfennig F., Bienvenu D., Gutzeit H.O., Volff J.N., Wenning M., Baroiller J.F. 2007. *Sox9a*, *Sox9b* and *Amh* are up-regulated in the gonads during natural and temperature-induced tilapia male differentiation. In Proceedings of the 8th International Symposium on Reproductive Physiology of Fish, June 3rd–8th, Saint-Malo, France.

Davey J.W., Blaxter M.L. 2010. RADSeq: next-generation population genetics. Briefings in Functional Genomics, 9(5-6):416-423.

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. Nature Reviews Genetics, 12:499-510.

de Almeida Toledo L.F., Foresti F. 2001. Morphologically differentiated sex chromosomes in neotropical freshwater fish. Genetica, 111(1-3):91-100.

de Verdal H., Rosario W., Vandeputte M., Muyalde N., Morissens P., Baroiller J.F., Chevassus B. 2014b. Response to selection for growth in an interspecific hybrid between *Oreochromis mossambicus* and *O. niloticus* in two distinct environments. Aquaculture, 430:159-165.

de Verdal H., Vandeputte M., Pepey E., Vidal M.O., Chatain B. 2014a. Individual growth monitoring of European sea bass larvae by image analysis and microsatellite genotyping. Aquaculture, 434:470-475.

Deines A.M., Bbole I., Katongo C., Feder J.L., Lodge D.M. 2014. Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. African Journal of Aquatic Science, 39(1):23-34.

Desjardins J.K., Fernald R.D. 2009. Fish sex: why so diverse? Current Opinion in Neurobiology, 19(6):648.

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*). Canadian Journal of Fisheries and Aquatic Sciences, 48(9):1606-1612.

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

Dominik S., Henshall J.M., Kube P.D., King H., Lien S., Kent M.P., Elliott N.G. 2010. Evaluation of an Atlantic salmon SNP chip as a genomic tool for the application in a Tasmanian Atlantic salmon (*Salmo salar*) breeding population. Aquaculture, 308(suppl. 1):S56-S61.

Donaldson E.M., Hunter G.A. 1982. Sex control in fish with particular reference to salmonids. Canadian Journal of Fisheries and Aquatic Sciences, 39:99-110.

Du Q.Y., Wang F.Y., Hua H.Y., Chang Z.J. 2007. Cloning and study of adult-tissue-specific expression of *Sox9* in *Cyprinus carpio*. Journal of Genetics, 86(2):85-91.

Dunham R.A. 2007. Comparison of six generations of selection, interspecific hybridization, intraspecific crossbreeding and gene transfer for growth improvement in ictalurid catfish. Aquaculture, 272:S252-S253.

Dunham R.A. 2011. Aquaculture and fisheries biotechnology: genetic approaches. CABI publishing, UK.

Eisbrenner W.D., Botwright N., Cook M., Davidson E.A., Dominik S., Elliott N.G., Henshall J., Jones S.L., Kube P.D., Lubieniecki K.P., Peng S., Davidson W.S. 2014. Evidence for multiple sex-determining loci in Tasmanian Atlantic salmon (*Salmo salar*). Heredity, 113(1):86-92.

Eknath A.E. 1995. Managing aquatic genetic resources, management example 4: the Nile tilapia. In: Conservation of Fish and Shellfish Resources: Managing Diversity (eds.

Thorpe J.E., Gall G.A., Lannan J.E., Nash C.E.). Academic Press, London (pp. 176-194).

Eknath A.E., Acosta B.O. 1998. Genetic improvement of farmed tilapias (GIFT) project: Final report, March 1988 to December 1997. The WorldFish Center Working Papers.

Eknath A.E., Macaranas J.M., Augustin L.Q., Velasco R.R., Ablan M.C.A., Pante M.J.R., Pullin R.S.V. 1991. Biochemical and morphometric approaches to characterise farmed tilapias. Naga, The ICLARM Quarterly, 14(2):7-9.

Ellegren H. 2001. Hens, cocks and avian sex determination. A quest for genes on Z or W? EMBO Reports, 2:192-196.

Ellsworth D.L., Rittenhouse K.D., Honeycutt R.L. 1993. Artifactual variation in randomly amplified polymorphic DNA banding patterns. BioTechniques, 14(2):214-217.

Embody G.C., Hayford C.O. 1925. The advantage of rearing brook trout fingerlings from selected breeders. Transactions of the American Fisheries Society, 55(1):135-148.

Eshel O., Shirak A., Dor L., Band M., Zak T., Markovich-Gordon M., Chalifa-Caspi V., Feldmesser E., Weller J.I., Seroussi E., Hulata G., Ron M. 2014. Identification of male-specific *Amh* duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). BMC Genomics, 15:774.

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

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

Eusèbe D.C., di Clemente N., Rey R., Pieau C., Vigier B., Josso N., Picard J.Y. 1996. Cloning and expression of the chick anti-Müllerian hormone gene. The Journal of Biological Chemistry, 271(9):4798-4804.

Everett M.V, Miller M.R., Seeb J.E. 2012. Meiotic maps of sockeye salmon derived from massively parallel DNA sequencing. BMC Genomics, 13:521.

Ezaz M.T., Harvey S.C., Boonphakdee C., Teale A.J., McAndrew B.J., Penman D.J. 2004. Isolation and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus* L.). Marine Biotechnology, 6(5):435-445.

Ezaz T., Stiglec R., Veyrunes F., Marshall Graves J.A. 2006. Relationships between vertebrate ZW and XY sex chromosome systems. Current Biology, 16(17):R736-R743.

Felip A., Martínez-Rodríguez G., Piferrer F., Carrillo M., Zanuy S. 2000. AFLP analysis confirms exclusive maternal genomic contribution of meiogynogenetic sea bass (*Dicentrarchus labrax* L.). Marine Biotechnology, 2(3):301-306.

Felip A., Young W.P., Wheeler P.A., Thorgaard G.H. 2005. An AFLP-based approach for the identification of sex-linked markers in rainbow trout (*Oncorhynchus mykiss*). Aquaculture, 247(1-4):35-43.

Ferguson-Smith M. 2007. The evolution of sex chromosomes and sex determination in vertebrates and the key role of *Dmrt1*. Sexual Development, 1(1):2-11.

Fernandino J.I., Hattori R.S., Shinoda T., Kimura H., Strobl-Mazzulla P.H., Strüssmann C.A., Somoza G.M. 2008. Dimorphic expression of *Dmrt1* and *Cyp19a1* (ovarian aromatase) during early gonadal development in pejerrey, *Odontesthes bonariensis*. Sexual Development, 2(6):316-324.

Ferreira I.A., Poletto A.B., Kocher T.D., Mota-Velasco J.C., Penman D.J., Martins C. 2010. Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. Cytogenetic and Genome Research, 129(4):314-322.

Ferretti L., Ramos-Onsins S.E., Pérez-Enciso M. 2013. Population genomics from pool sequencing. Molecular Ecology, 22(22):5561-5576.

Firmat C., Alibert P., Losseau M., Baroiller J.F., Schliewen U.K. 2013. Successive invasion mediated interspecific hybridizations and population structure in the endangered cichlid *Oreochromis mossambicus*. Plos One, 8(5):e63880.

Fishelson L. 1966. Cichlid fish of the genus *Tilapia* in Israel. Bamidge, 18:67-80.

Fishelson L. 1970. Protogynous sex reversal in the fish *Anthias squamipinnis* (Teleostei, Anthiidae) regulated by the presence or absence of a male fish. Nature, 227:90-91.

Fjalestad K.T., Gjedrem T., Carr W.H., Sweeney J.N. 1997. Final report: the shrimp breeding program. Selective breeding of *Penaeus vannamei*. AKVAFORSK, Report no. 17/97, (p. 85).

Ford C.E., Jones K.W, Polani P.E., De Almeida J.C., Briggs J.H. 1959. A sex-chromosome anomaly in a case of gonadal dysgenesis (Turner's syndrome). The Lancet, 273(7075):711-713.

Francis R.C. 1984. The effects of bidirectional selection for social dominance on agonistic behavior and sex ratios in the paradise fish (*Macropodus opercularis*). Behaviour, 90(1):25-45.

Froschauer A., Körting 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(2):247-254.

Futschik A., Schlötterer C. 2010. The next generation of molecular markers from massively parallel sequencing of pooled DNA samples. Genetics, 186(1):207-218.

Gautier M., Foucaud J., Gharbi K., Cézard T., Galan M., Loiseau A., Thomson M., Pudlo P., Kerdelhué C., Estoup A. 2013b. Estimation of population allele frequencies from next-generation sequencing data: pool-versus individual-based genotyping. Molecular Ecology, 22(14):3766-3779.

Gautier M., Gharbi K., Cézard T., Foucaud J., Kerdelhué C., Pudlo P., Cornuet J.M., Estoup A. 2013a. The effect of RAD allele dropout on the estimation of genetic variation within and between populations. Molecular Ecology, 22(11):3165-3178.

Gennotte V., Mafwila Kinkela P., Ulysse B., Akian Djétouan D., Bere Sompagnimdi F., Tomson T., Mélard C., Rougeot C. 2015. Brief exposure of embryos to steroids or aromatase inhibitor induces sex reversal in Nile tilapia (*Oreochromis niloticus*). Journal of Experimental Zoology, 323(1):31-38.

Gilbey J., Verspoor E., McLay A., Houlihan D. 2004. A microsatellite linkage map for Atlantic salmon (*Salmo salar*). Animal Genetics, 35(2):98-105.

Gilling C.J., Skibinski D.O.F., Beardmore J.A. 1996. Sex reversal of tilapia fry by immersion in water containing estrogens. In ICLARM Conference Proceedings, Makati City (Philippines). ICLARM, Makati City, Philippines (pp. 314-319).

Giovannoni J.J., Wing R.A., Ganal M.W., Tanksley S.D. 1991. Isolation of molecular markers from specific chromosomal intervals using DNA pools from existing mapping populations. Nucleic Acids Research, 19(23):6553-6568.

Gitterle T., Johansen H., Erazo C., Lozano C., Cock J., Salazar M., Rye M. 2007. Response to multi-trait selection for harvest weight, overall survival, and resistance to white spot syndrome virus (WSSV) in *Penaeus* (*Litopenaeus*) *vannamei*. Aquaculture, 272:S262.

Gjedrem T. 1985. Improvement of productivity through breeding schemes. GeoJournal, 10(3):233-241.

Gjedrem T. 2010. The first family-based breeding program in aquaculture. Reviews in Aquaculture, 2(1):2-15.

Gjedrem T., Robinson N., Rye M. 2012. The importance of selective breeding in aquaculture to meet future demands for animal protein: a review. Aquaculture, 350-353:117-129.

Godwin J., Crews D., Warner R.R. 1996. Behavioural sex change in the absence of gonads in a coral reef fish. Proceedings of the Royal Society B: Biological Sciences, 263(1377):1683-1688.

Golan M., Levavi-Sivan B. 2014. Artificial masculinization in tilapia involves androgen receptor activation. General and Comparative Endocrinology, 207:50-55.

Göth A., Booth D.T. 2005. Temperature-dependent sex ratio in a bird. Biology Letters, 1(1):31-33.

Goudie C.A., Khan G., Parker N. 1985. Gynogenesis and sex manipulation with evidence for female homogameity in channel catfish (*Ictalurus punctatus*). In: Annual Progress Report 1 October–30 September. USFWS, Southeastern Fish Culture Laboratory, Marion, Alabama.

Goudie C.A., Redner B.D., Simco B.A., Davis K.B. 1983. Feminization of channel catfish by oral administration of steroid sex hormones. Transactions of the American Fisheries Society, 112:670-672.

Gregg R.E., Howard J.H., Shonhiwa F. 1998. Introgressive hybridization of tilapias in Zimbabwe. Journal of Fish Biology, 52(1):1-10.

Griffiths R., Orr K.L., Adam A., Barber I. 2000. DNA sex identification in the Three-spined stickleback. Journal of Fish Biology, 57(5):1331-1334.

Groenen M.A.M., Megens H.J., Zare Y., Warren W.C., Hillier L.W., Crooijmans R.P.M.A., Vereijken A., Okimoto R., Muir W.M., Cheng H.H. 2011. The development and characterization of a 60K SNP chip for chicken. BMC Genomics, 12:274.

Guerrero III R.D. 1975. Use of androgens for the production of all-male *Tilapia aurea* (Steindachner). Transactions of the American Fisheries Society, 104(2):342-348.

Guerrero III R.D., Shelton W.L. 1974. An aceto-carmine squash method for sexing juvenile fishes. The Progressive Fish-Culturist, 36(1):56-56.

Guiguen Y., Baroiller J.F., Ricordel M.J., Iseki K., McMeel O.M., Martin S.A.M., Fostier A. 1999. Involvement of estrogens in the process of sex differentiation in two fish species: the rainbow trout (*Oncorhynchus mykiss*) and a tilapia (*Oreochromis niloticus*). Molecular Reproduction and Development, 54(2):154-162.

Guiguen Y., Fostier A., Piferrer F., Chang C.F. 2010. Ovarian aromatase and estrogens: a pivotal role for gonadal sex differentiation and sex change in fish. General and Comparative Endocrinology, 165(3):352-366.

Guo Y., Cheng H., Huang X., Gao S., Yu H., Zhou R. 2005. Gene structure, multiple alternative splicing, and expression in gonads of Zebrafish *Dmrt1*. Biochemical and Biophysical Research Communications, 330(3):950-957.

Gutierrez A.P., Lubieniecki K.P., Davidson E.A., Lien S., Kent M.P., Fukui S., Withler R.E., Swift B., Davidson W.S. 2012. Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. Aquaculture, 358:61-70.

Hale M.C., Jackson J.R., DeWoody J.A. 2010. Discovery and evaluation of candidate sex-determining genes and xenobiotics in the gonads of lake sturgeon (*Acipenser fulvescens*). Genetica, 138(7):745-756.

Halm S., Rocha A., Miura T., Prat F., Zanuy S. 2007. Anti-Müllerian hormone (*Amh*/*Amh*) in the European sea bass: Its gene structure, regulatory elements, and the expression of alternatively-spliced isoforms. Gene, 388:148-158.

Hamdoon N.T., Ibrahim F., Kelany A.M., Elshazly H.F., Zayed A.E. 2013. Hormonal sex reversal in *Oreochromis niloticus* by oral administration of diethylstilbesterole. Life Science Journal, 10(2):123-128.

Hassanien H.A., Elnady M., Obeida A., Itriby H. 2004. Genetic diversity of Nile tilapia populations revealed by randomly amplified polymorphic DNA (RAPD). Aquaculture Research, 35(6):587-593.

Hattori R.S., Murai Y., Oura M., Masuda S., Majhi S.K., Sakamoto T., Fernandino J.I., Somoza G.M., Yokota M., Strüssmann C.A. 2012. A Y-linked anti-Müllerian hormone duplication takes over a critical role in sex determination. Proceedings of the National Academy of Sciences of the United States of America, 109(8):2955-2959.

Hawkes L.A., Broderick A.C., Godfrey M.H., Godley B.J. 2009. Climate change and marine turtles. Endangered Species Research, 7(2):137-154.

Hayes T.B. 1998. Sex determination and primary sex differentiation in amphibians: genetic and developmental mechanisms. The Journal of Experimental Zoology, 281(5):373-399.

Hemmer-Hansen J., Therkildsen N.O., Pujolar J.M. 2014. Population genomics of marine fishes: next-generation prospects and challenges. The Biological Bulletin, 227(2):117-132.

Henley W.F., Grobler P.J., Neves R.J. 2006. Non-invasive method to obtain DNA from freshwater mussels (Bivalvia: Unionidae). Journal of Shellfish Research, 25(3):975-977.

Herpin A., Schartl M. 2011. *Dmrt1* genes at the crossroads: a widespread and central class of sexual development factors in fish. The FEBS Journal, 278(7):1010-1019.

Hetzel D.J.S., Crocos P.J., Davis G.P., Moore S.S., Preston N.C. 2000. Response to selection and heritability for growth in the Kuruma prawn, *Penaeus japanicus*. Aquaculture, 181:215–223.

Hickling C.F. 1960. The Malacca tilapia hybrids. Journal of Genetics, 57(1):1-10.

Hill W.G. 1982. Dominance and epistasis as components of heterosis. Journal of Animal Breeding and Genetics, 99(1-4):161-168.

Hillis D.M., Moritz C., Mable B.K. 1996. Molecular systematics, Sinauer Associates Sunderland, Massachusetts USA.

Hisar O., Sönmez A.Y., Hisar Ş.A., Budak H., Gültepe N. 2013. The sexually dimorphic adipose fin is an androgen target tissue in the brown trout (*Salmo trutta fario*). Fish Physiology and Biochemistry, 39:233-241.

Hohenlohe P.A., Amish S.J., Catchen J.M., Allendorf F.W., Luikart G. 2011. Next-generation RAD sequencing identifies thousands of SNPs for assessing hybridization between rainbow and westslope cutthroat trout. Molecular Ecology Resources, 11(Suppl. 1):117-122.

Hohenlohe 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 Genetics, 6(2):e1000862.

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

Hoolihan J.P., Perez N.F., Faugue R.M., Bernard A.M., Horn R.L., Snodgrass D., Schultz D.R. 2009. Surface mucous as a source of genomic DNA from Atlantic billfishes (Istiophoridae) and swordfish (Xiphiidae). Fishery Bulletin, 107(3):339-342.

Hopkins K.D., Shelton W.L., Engle C.R. 1979. Estrogen sex-reversal of *Tilapia aurea*. Aquaculture, 18:263-268.

Hourigan T.F., Nakamura M., Nagahama Y., Yamauchi K., Grau E.G. 1991. Histology, ultrastructure, and in vitro steroidogenesis of the testes of two male phenotypes of the protogynous fish, *Thalassoma duperrey* (labridae). General and Comparative Endocrinology, 83(2):193-217.

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.

Hovda M.B., Sivertsvik M., Lunestad B.T., Lorentzen G., Rosnes J.T. 2007. Characterisation of the dominant bacterial population in modified atmosphere packaged farmed halibut (*Hippoglossus hippoglossus*) based on 16S rDNA-DGGE. Food Microbiology, 24(4):362-371.

Huang S., Deng L., Guan M., Li J., Lu K., Wang H., Fu D., Mason A.S., Liu S., Hua W. 2013. Identification of genome-wide single nucleotide polymorphisms in allopolyploid crop *Brassica napus*. BMC Genomics, 14:717.

Hulata G., Wohlfarth G., Rothbard S. 1983. Progeny-testing selection of tilapia broodstocks producing all-male hybrid progenies-preliminary results. Aquaculture, 33:263-268.

Hussain M.G., Islam M.S., Hossain M.A., Wahid M.I., Kohinoor A.H.M., Dey M.M., Mazid M.A. 2002. Stock improvement of silver carp (*Barbodes gonionotus* Bleeker) through several generations of genetic selection. Aquaculture, 204:469-480.

Huyhn K., Renfree M.B., Graves J.A., Pask A.J. 2011. ATRX has a critical and conserved role in mammalian sexual differentiation. BMC Developmental Biology, 11:39.

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(2):333-41.

Iturra P., Medrano J.F., Bagley M., Lam N., Vergara N., Marin J.C. 1997. Identification of sex chromosome molecular markers using RAPDs and fluorescent in situ hybridization in rainbow trout. Genetica, 101(3):209-213.

Jacobs P.A., Strong J.A. 1959. A case of human intersexuality having a possible XXY sex-determining mechanism. Nature, 183:302-303.

Jakob S., Lovell-Badge R. 2011. Sex determination and the control of *Sox9* expression in mammals. The FEBS Journal, 278(7):1002-1009.

Jalabert B., Billard R., Chevassus B. 1975. Preliminary experiments on sex control in trout: production of sterile fishes and simultaneous self-fertilizable hermaphrodites. Annales de Biologie Animale Biochimie Biophysique, 15:19-28.

Jamin S.P., Arango N.A., Mishina Y., Hanks M.C., Behringer R.R. 2003. Genetic studies of the *Amh*/mis signaling pathway for Müllerian duct regression. Molecular and Cellular Endocrinology, 211:15-19.

Janzen F.J., Phillips P.C. 2006. Exploring the evolution of environmental sex determination, especially in reptiles. Journal of Evolutionary Biology, 19(6):1775-1784.

Jo P.G., An K.W., Kim N.N., Choi Y.K., Cho S.H., Min B.H., Lim H.K., Choi C.Y. 2007. Induced expression of doublesex-and mab-3-related transcription factor-1 (*Dmrt1*) mRNA by testosterone in the olive flounder, *Paralichthys olivaceus* ovary. Journal of Aquaculture, 20(3):199-202.

Johnsen H., Seppola M., Torgersen J.S., Delghandi M., Andersen Ø. 2010. Sexually dimorphic expression of *Dmrt1* in immature and mature Atlantic cod (*Gadus morhua* L.). Comparative Biochemistry and Physiology Part B: Biochemistry & Molecular Biology, 156(3):197-205.

Johnsen H., Tveiten H., Torgersen J.S., Andersen Ø. 2013. Divergent and sex-dimorphic expression of the paralogs of the *Sox9-Amh-Cyp19a1* regulatory cascade in developing and adult atlantic cod (*Gadus morhua* L.). Molecular Reproduction and Development, 80(5):358-370.

Josso N. 2011. Biology and genetics of anti-Müllerian hormone. In: Hormonal and genetic basis of sexual differentiation disorders and hot topics in Endocrinology (eds. New M.I., Simpson J.L.). Proceedings of the 2nd World Conference, Springer New York (pp. 83-85).

Josso N., di Clemente N., Gouédard L. 2001. Anti-Müllerian hormone and its receptors. Molecular and Cellular Endocrinology, 179:25-32.

Kageyama D., Narita S., Watanabe M. 2012. Insect sex determination manipulated by their endosymbionts: incidences, mechanisms and implications. Insects, 3(1):161-199.

Kai W., Kikuchi K., Fujita M., Suetake H., Fujiwara A., Yoshiura Y., Ototake M., Venkatesh B., Miyaki K., Suzuki Y. 2005. A genetic linkage map for the tiger pufferfish, *Takifugu rubripes*. Genetics, 171(1):227-238.

Kamiya T., Kai W., Tasumi S., Oka A., Matsunaga T., Mizuno N., Fujita M., Suetake H., Suzuki S., Hosoya S., Tohari S., Brenner S., Miyadai T., Venkatesh B., Suzuki Y., Kikuchi K. 2012. A trans-species missense SNP in *Amhr2* is associated with sex determination in the tiger pufferfish, *Takifugu rubripes* (fugu). Plos Genetics, 8(7):e1002798.

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.

Kazeto Y., Tosaka R., Matsubara H., Ijiri S., Adachi S. 2011. Ovarian steroidogenesis and the role of sex steroid hormones on ovarian growth and maturation of the Japanese eel. The Journal of Steroid Biochemistry and Molecular Biology, 127:149-154.

Keyvanshokooh S., Pourkazemi M., Kalbassi M.R. 2007. The RAPD technique failed to identify sex-specific sequences in beluga (*Huso huso*). Journal of Applied Ichthyology, 23(1):1-2.

Kikuchi K., Kai W., Hosokawa A., Mizuno N., Suetake H., Asahina K., Suzuki Y. 2007. The sex-determining locus in the tiger pufferfish, *Takifugu rubripes*. Genetics, 175(4):2039-2042.

Kincaid H.L., Bridges W.R., von Limbach B. 1977. Three generations of selection for growth rate in fall spawning rainbow trout. Transactions of the American Fisheries Society, 106:621-629.

Kirov G., Williams N., Sham P., Craddock N., Owen M.J. 2000. Pooled genotyping of microsatellite markers in parent-offspring trios. Genome Research, 10:105-115.

Kobayashi H., Iwamatsu T. 2005. Sex reversal in the medaka *Oryzias latipes* by brief exposure of early embryos to estradiol-17β. Zoological Science, 22(10):1163-1167.

Kobayashi T., Kajiura-Kobayashi H., Guan G., Nagahama Y. 2008. Sexual dimorphic expression of *Dmrt1* and *Sox9a* during gonadal differentiation and hormone-induced sex reversal in the teleost fish Nile tilapia (*Oreochromis niloticus*). Developmental Dynamics, 237(1):297-306.

Kobayashi T., Kajiura-Kobayashi H., Nagahama Y. 2003. Induction of XY sex reversal by estrogen involves altered gene expression in a teleost, tilapia. Cytogenetic and Genome Research, 101:289-294.

Kobayashi T., Matsuda M., Kajiura-Kobayashi H., Suzuki A., Saito N., Nakamoto M., Shibata N., Nagahama Y. 2004. Two DM domain genes, *Dmy* and *Dmrt1*, involved in testicular differentiation and development in the Medaka, *Oryzias latipes*. Developmental Dynamics, 231(3):518-526.

Kobayashi Y., Horiguchi R., Miura S., Nakamura M. 2010. Sex- and tissue-specific expression of P450 aromatase (*Cyp19a1a*) in the yellowtail clownfish, *Amphiprion clarkii*. Comparative Biochemistry & Physiology Part A: Molecular & Integrative Physiology, 155(2):237-244.

Kobayashi Y., Nagahama Y., Nakamura M. 2013. Diversity and plasticity of sex determination and differentiation in fishes. Sexual Development, 7:115-125.

Kocher T.D., Lee W.J., Sobolewska H., Penman D.J., McAndrew B.J. 1998. A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). Genetics, 148(3):1225-1232.

Kondo M., Hornung U., Nanda I., Imai S., Sasaki T., Shimizu A., Asakawa S., Hori H., Schmid M., Shimizu N., Schartl M. 2006. Genomic organization of the sex-determining and adjacent regions of the sex chromosomes of Medaka. Genome Research, 16(7):815-826.

Koopman P. 2001. The genetics and biology of vertebrate sex determination. Cell, 105(7):843-847.

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

Koopman P., Schepers G., Brenner S., Venkatesh B. 2004. Origin and diversity of the *Sox* transcription factor gene family: genome-wide analysis in *Fugu rubripes*. Gene, 328:177-186.

Kopp A. 2012. *Dmrt* genes in the development and evolution of sexual dimorphism. Trends in Genetics, 28(4):175-184.

Kovács B., Egedi S., Bártfai R., Orbán L. 2000. Male-specific DNA markers from African catfish (*Clarias gariepinus*). Genetica, 110(3):267-276.

Kucuktas H., Wang S., Li P., He C., Xu P., Sha Z., Liu H., Jiang Y., Baoprasertkul P., Somridhivej B., Wang Y., Abernathy J., Guo X., Liu L., Muir W., Liu Z. 2009. Construction of genetic linkage maps and comparative genome analysis of catfish using gene-associated markers. Genetics, 181(4):1649-1660.

Kuzema A.I. 1971. The Ukrainian breeds of carp. Seminar/study tour in USSR on genetic selection and hybridization of cultivated fishes. FAO, Rome (pp. 228–232).

Kwon J.Y., McAndrew B.J., Penman D.J. 2002. Treatment with an aromatase inhibitor suppresses high-temperature feminization of genetic male (YY) Nile tilapia. Journal of Fish Biology, 60(3):625-636.

Lander E.S., Botstein D. 1989. Mapping Mendelian factors underlying quantitative traits using RFLP linkage maps. Genetics, 121:185-199.

Langdon C., Evans F., Jacobson D., Blouin M. 2003. Yields of cultured pacific oysters, *Crassostrea gigas* Thunberg, improved after one generation of selection. Aquaculture, 220:227-244.

Langmead B. 2010. Aligning short sequencing reads with Bowtie. Current Protocols in Bioinformatics, 32:11.7.1-11.7.14.

Langmead B., Salzberg S.L. 2012. Fast gapped-read alignment with Bowtie 2. Nature Methods, 9:357-359.

Larsen A., Tao Z., Bullard S.A., Arias C.R. 2013. Diversity of the skin microbiota of fishes: evidence for host species specificity. FEMS Microbiology Ecology, 85:483-494.

Le Vin A.L., Adam A., Tedder A., Arnold K.E., Mable B.K. 2011. Validation of swabs as a non-destructive and relatively non-invasive DNA sampling method in fish. Molecular Ecology Resources, 11(1):107-109.

Lee B.Y., Coutanceau J.P., Ozouf-Costaz C., D'Cotta H., Baroiller J.F., Kocher T.D. 2011. Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). Marine Biotechnology, 13(3):557-562.

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

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 genetic linkage map of tilapia (*Oreochromis* spp.). Genetics, 170(1):237-244.

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. Animal Genetics, 34(5):379-383.

Li M., Sun Y., Zhao J., Shi H., Zeng S., Ye K., Jiang D., Zhou L., Sun L., Tao W., Nagahama Y., Kocher T.D. 2015. A tandem duplicate of anti-Müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. Plos Genetics, 11:e1005678.

Lieber L., Berrow S., Johnston E., Hall G., Hall J., Gubili C., Sims D.W., Jones C.S., Noble L.R. 2013. Mucus: aiding elasmobranch conservation through non-invasive genetic sampling. Endangered Species Research, 21:215-222.

Lien S., Gidskehaug L., Moen T., Hayes B.J., Berg P.R., Davidson W.S., Omholt S.W., Kent M.P. 2011. A dense SNP-based linkage map for Atlantic salmon (*Salmo salar*) reveals extended chromosome homeologies and striking differences in sex-specific recombination patterns. BMC Genomics, 12:615.

Litt M., Luty J.A. 1989. A hypervariable microsatellite revealed by in vitro amplification of a dinucleotide repeat within the cardiac muscle actin gene. American Journal of Human Genetics, 44(3):397-401.

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.

Liu Q., Goudie C.A., Simco B.A., Davis K.B. 1996. Sex-linkage of glucosephosphate isomerase-B and mapping of the sex-determining gene in channel catfish. Cytogenetic and Genome Research, 73(4):282-285.

Liu Z.H., Zhang Y.G., Wang D.S. 2010. Studies on feminization, sex determination, and differentiation of the Southern catfish, *Silurus meridionalis*-a review. Fish Physiology and Biochemistry, 36(2):223-235.

Liu Z.J., Cordes J.F. 2004. DNA marker technologies and their applications in aquaculture genetics. Aquaculture, 238:1-37.

Livaja M., Wang Y., Wieckhorst S., Haseneyer G., Seidel M., Hahn V., Knapp S.J., Taudien S., Schön C.C., Bauer E. 2013. BSTA: a targeted approach combines bulked segregant analysis with next-generation sequencing and *de novo* transcriptome assembly for SNP discovery in sunflower. BMC Genomics, 14:628.

Livia L., Antonella P., Hovirag L., Mauro N., Panara F. 2006. A nondestructive, rapid, reliable and inexpensive method to sample, store and extract high-quality DNA from fish body mucus and buccal cells. Molecular Ecology Notes, 6:257-260.

Luca F., Hudson R.R., Witonsky D.B., Di Rienzo A. 2011. A reduced representation approach to population genetic analyses and applications to human evolution. Genome Research, 21(7):1087-1098.

Luckenbach J.A., Borski R.J., Daniels H.V., Godwin J. 2009. Sex determination in flatfishes: mechanisms and environmental influences. Seminars in Cell & Developmental Biology, 20(3):256-263.

Luckenbach J.A., Godwin J., Daniels H.V., Borski R.J. 2003. Gonadal differentiation and effects of temperature on sex determination in Southern flounder (*Paralichthys lethostigma*). Aquaculture, 216:315-327.

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*). Sexual Development, 6(5):247-256.

Ma K., Liao M., Liu F., Ye B., Sun F., Yue G.H. 2016. Charactering the *Zfand3* gene mapped in the sex-determining locus in hybrid tilapia (*Oreochromis* spp.). Scientific Reports, 6:25471.

Macaranas J.M., Taniguchi N., Pante M.J.R., Capili J.B., Pullin R.S.V. 1986. Electrophoretic evidence for extensive hybrid gene introgression into commercial *Oreochromis niloticus* (L.) stocks in the Philippines. Aquaculture Research, 17(4):249-258.

Mair G.C., Abucay J.S., Abella T.A., Beardmore J.A., Skibinski D.O.F. 1997. Genetic manipulation of sex ratio for the large-scale production of all-male tilapia, *Oreochromis niloticus*. Canadian Journal of Fisheries and Aquatic Sciences, 54(2):396-404.

Mair G.C., Abucay J.S., Beardmore J.A., Skibinski D.O.F. 1995. Growth performance trials of genetically male tilapia (GMT) derived from YY-males in *Oreochromis niloti*cus L.: on station comparisons with mixed sex and sex reversed male populations. Aquaculture, 137:313-323.

Mair G.C., Scott A.G., Penman D.J., Beardmore J.A., Skibinski D.O. 1991. Sex determination in the genus *Oreochromis*: 1. Sex reversal, gynogenesis and triploidy in *O. niloticus* (L.). Theoretical and Applied Genetics, 82(2):144-152.

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(3):175-188.

Mank J.E., Promislow D.E.L., Avise J.C. 2006. Evolution of alternative sex determining mechanisms in teleost fishes. Biological Journal of the Linnean Society, 87(1):83-93.

Marchand O., Govoroun M., D'Cotta H., McMeel O., Lareyre J.J., Bernot A., Laudet V., Guiguen Y. 2000. *Dmrt1* expression during gonadal differentiation and spermatogenesis in the rainbow trout, *Oncorhynchus mykiss*. Biochimica et Biophysica Acta (BBA) - Gene Structure and Expression, 1493:180-187.

Mardis E.R. 2008a. Next-generation DNA sequencing methods. Annual Review of Genomics and Human Genetics, 9:387-402.

Mardis E.R. 2008b. The impact of next-generation sequencing technology on genetics. Trends in Genetics, 24(3):133-141.

Matsuda M., Nagahama Y., Shinomiya A., Sato T., Matsuda C., Kobayashi T., Morrey C.E., Shibata N., Asakawa S., Shimizu N., Hori H., Hamaguchi S., Sakaizumi M. 2002. *Dmy* is a Y-specific DM-domain gene required for male development in the Medaka fish. Nature, 417:559-563.

Matsuda M., Sato T., Toyazaki Y., Nagahama Y., Hamaguchi S., Sakaizumi M. 2003. *Oryzias curvinotus* has *Dmy*, a gene that is required for male development in the medaka, *O. latipes*. Zoological Science, 20(2):159-161.

Mawaribuchi S., Yoshimoto S., Ohashi S., Takamatsu N., Ito M. 2012. Molecular evolution of vertebrate sex-determining genes. Chromosome Research, 20(1):139-151.

McAndrew B.J. 2000. Evolution, phylogenetic relationships and biogeography. In: Tilapias biology and exploitation (eds. Beveridge M.C.M., McAndrew B.J.). Kluwer Academic Publishers (pp. 1-32).

McAndrew B.J., Roubal F.R., Roberts R.J., Bullock A.M., McEwen I.M. 1988. The genetics and histology of red, blond and associated colour variants in *Oreochromis niloticus*. Genetica, 76(2):127-137.

McClung C.E. 1902. The accessory chromosome-sex determinant? The Biological Bulletin, 3(1-2):43-84.

Meissner H., Fevolden S.E., Amundsen P.A., Præbel K. 2013. Evaluation of three methods for high throughput extraction of DNA from challenging fish tissues. Conservation Genetics Resources, 5:733-735.

Melchers W., Van den Brule A., Walboomers J., de Bruin M., Burger M., Herbrink P., Meijer C., Lindeman J., Quint W. 1989. Increased detection rate of human papillomavirus in cervical scrapes by the polymerase chain reaction as compared to modified FISH and southern blot analysis. Journal of Medical Virology, 27:329-335.

Meriwether F.H., Shelton W.L. 1981. Observation on aquarium spawning of estrogen-treated and untreated tilapia. Proceedings Annual Conference Southeastern Association of Fish and Wildlife Agencies, 34:81-87.

Metzker M.L. 2010. Sequencing technologies - the next generation. Nature Reviews Genetics, 1:31-46.

Meuwissen T.H.E., Hayes B.J., Goddard M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics, 157(4):1819-1829.

Michelmore R.W., Paran I., Kesseli R.V. 1991. Identification of markers linked to disease-resistance genes by bulked segregant analysis: a rapid method to detect markers in specific genomic regions by using segregating populations. Proceedings of the National Academy of Sciences of the United States of America, 88(21):9828-9832.

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. Molecular Ecology, 21(2):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 Research, 17:240-248.

Mirimin L., O'Keeffe D., Ruggiero A., Bolton-Warberg M., Vartia S., Fitzgerald R. 2011. A quick, least-invasive, inexpensive and reliable method for sampling *Gadus morhua* postlarvae for genetic analysis. Journal of Fish Biology, 79:801-805.

Miura T., Miura C., Konda Y., Yamauchi K. 2002. Spermatogenesis-preventing substance in Japanese eel. Development, 129:2689-2697.

Modi W.S., Crews D. 2005. Sex chromosomes and sex determination in reptiles: Commentary. Current Opinion in Genetics & Development, 15(6):660-665.

Moralee R.D., van der Bank F.H., van der Waal B.C.W. 2000. Biochemical genetic markers to identify hybrids between the endemic *Oreochromis mossambicus* and the alien species, *O. niloticus* (Pisces: Cichlidae). Water SA, 26(2):263-268.

Morizot D.C., Slaugenhaupt S.A., Kallman K.D., Chakravarti A. 1991. Genetic linkage map of fishes of the genus *Xiphophorus* (Teleostei: Poeciliidae). Genetics, 127(2):399-410.

Morrison C.M., Miyake T., Wright Jr.J.R. 2001. Histological study of the development of the embryo and early larva of *Oreochromis niloticus*. (Pisces: Cichlidae). Journal of Morphology, 247:172-195.

Müller-Belecke A., Hörstgen-Schwark G. 1995. Sex determination in tilapia (*Oreochromis niloticus*) sex ratios in homozygous gynogenetic progeny and their offspring. Aquaculture, 137:57-65.

Murphy M.W., Sarver A.L., Rice D., Hatzi K., Ye K., Melnick A., Heckert L.L., Zarkower D., Bardwell V.J. 2010. Genome-wide analysis of DNA binding and transcriptional regulation by the mammalian doublesex homolog *Dmrt1* in the juvenile testis. Proceedings of the National Academy of Sciences of the United States of America, 107(30):13360-13365.

Mwanja W., Kaufman L. 1995. A note on recent advances in the genetic characterization of tilapia stocks in Lake Victoria region. African Journal of Tropical Hydrobiology and Fisheries, 6:51-53.

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(3):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(1):163-170.

Nagl S., Tichy H., Mayer W.E., Samonte I.E., McAndrew B.J., Klein J. 2001. Classification and phylogenetic relationships of African tilapiine fishes inferred from mitochondrial DNA sequences. Molecular Phylogenetics and Evolution, 20(3):361-374.

Naish K.A., Hard J.J. 2008. Bridging the gap between the genotype and the phenotype: linking genetic variation, selection and adaptation in fishes. Fish and Fisheries, 9(4):396-422.

Nakamoto M., Suzuki A., Matsuda M., Nagahama Y., Shibata N. 2005. Testicular type *Sox9* is not involved in sex determination but might be in the development of testicular structures in the Medaka, *Oryzias latipes*. Biochemical and Biophysical Research Communications, 333(3):729-736.

Nakamura M. 2009. Sex determination in amphibians. Seminars in Cell & Developmental Biology, 20(3):271-282.

Nakamura M., Kobayashi T., Chang X.T., Nagahama Y. 1998. Gonadal sex differentiation in teleost fish. Journal of Experimental Zoology, 281(5):362-372.

Nakamura M., Kobayashi Y., Miura S., Alam M.A., Bhandari R.K. 2005. Sex change in coral reef fish. Fish Physiology and Biochemistry, 31:117-122.

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*. Proceedings of the National Academy of Sciences of the United States of America, 99(18):11778-11783.

Nanda I., Shan Z., Schartl M., Burt D.W., Koehler M., Nothwang H.G., Grützner F., Paton I.R., Windsor D., Dunn I., Engel W., Staeheli P., Mizuno S., Haaf T., Schmid M. 1999. 300 million years of conserved synteny between chicken Z and human chromosome 9. Nature Genetics, 21(3):258-259.

Naruse K., Fukamachi S., Mitani H., Kondo M., Matsuoka T., Kondo S., Hanamura N., Morita Y., Hasegawa K., Nishigaki R., Shimada A., Wada H., Kusakabe T., Suzuki N., Kinoshita M., Kanamori A., Terado T., Kimura H., Nonaka M., Shima A. 2000. A detailed linkage map of Medaka, *Oryzias latipes*: comparative genomics and genome evolution. Genetics, 154(4):1773-1784.

Navarro-Martín L., Viñas J., Ribas L., Díaz N., Gutiérrez A., Di Croce L., Piferrer F. 2011. DNA methylation of the gonadal aromatase (*Cyp19a*) promoter is involved in

temperature-dependent sex ratio shifts in the European seabass. Plos Genetics, 7(12):e1002447.

Neira R. 2010. Breeding in aquaculture species: genetic improvement programs in developing countries. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany (p. 8).

Nell J.A., Hand R.E. 2003. Evaluation of the progeny of second-generation Sydney rock oyster *Saccostrea glomerata* (Gold 1850) breeding lines for resistance to QX disease *Marteilia sydneyi*. Aquaculture, 228:27-35.

Nelson J.S. 2006. Fishes of the World. John Wiley & Sons.

Newkirk G.F., Haley L.E. 1983. Selecting for growth rate in the European oyster, *Ostrea edulis*: response of second generation group. Aquaculture, 33:149-155.

Nishioka M., Miura I., Saitoh K. 1993. Sex Chromosomes of *Rana rugosa* with special reference to local differences in sex-determining mechanism. Scientific Report of the Laboratory for Amphibian Biology, 12:55-81.

Ogawa Y., Taketomi Y., Murakami M., Tsujimoto M., Yanoshita R. 2013. Small RNA transcriptomes of two types of exosomes in human whole saliva determined by next generation sequencing. Biological & Pharmaceutical Bulletin, 36:66-75.

Okada E., Yoshimoto S., Ikeda N., Kanda H., Tamura K., Shiba T., Takamatsu N., Ito M. 2009. *Xenopus* W-linked *Dmw* induces *Foxl2* and *Cyp19* expression during ovary formation. Sexual Development, 3(1):38-42.

Olesen I., Gjedrem T., Bentsen H.B., Gjerde B., Rye M. 2003. Breeding programs for sustainable aquaculture. Journal of Applied Aquaculture, 13:179-204.

Ospina-Álvarez N., Piferrer F. 2008. Temperature-dependent sex determination in fish revisited: prevalence, a single sex ratio response pattern, and possible effects of climate change. Plos One, 3(7):e2837.

Ozaki A., Araki K., Okamoto H., Okauchi M., Mushiake K., Yoshida K., Tsuzaki T., Fuji K., Sakamoto T., Okamoto N. 2012. Progress of DNA marker-assisted breeding in maricultured finfish. Bulletin of Fisheries Research Agency, 35:31-37.

Page D.C., Mosher R., Simpson E.M., Fisher E.M.C., Mardon G., Pollack J., McGillivray B., de la Chapelle A., Brown L.G. 1987. The sex-determining region of the human Y chromosome encodes a finger protein. Cell, 51(6):1091-1104.

Painter T.S. 1923. Studies in mammalian spermatogenesis. II. The spermatogenesis of man. Journal of Experimental Zoology, 37(3):291-336.

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, 14:566.

Palaiokostas C., Bekaert M., Khan M.G.Q., Taggart J.B., Gharbi K., McAndrew B.J., Penman D.J. 2013a. Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) using RAD sequencing. Plos One, 8(7):e68389.

Palaiokostas C., Bekaert M., Khan M.G.Q., Taggart J.B., Gharbi K., McAndrew B.J., Penman D.J. 2015. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). BMC Genomics, 16:171.

Palti Y., Parsons J.E., Thorgaard G.H. 1999. Identification of candidate DNA markers associated with IHN virus resistance in backcrosses of Rainbow (*Oncorhynchus mykiss*) and Cutthroat trout (*O. clarki*). Aquaculture, 173:81-94.

Pandey S.C., Nandi G.C. 2014. TSD based framework for mining the induction rules. Journal of Computational Science, 5(2):184-195.

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

Parnell N.F., Streelman J.T. 2013. Genetic interactions controlling sex and color establish the potential for sexual conflict in Lake Malawi cichlid fishes. Heredity, 110:239-246.

Patiño R., Davis K.B., Schoore J.E., Uguz C., Strüssmann C.A., Parker N.C., Simco B.A., Goudie C.A. 1996. Sex differentiation of channel catfish gonads: normal development and effects of temperature. Journal of Experimental Zoology, 276(3):209-218.

Peichel C.L., Ross J.A., Matson C.K., Dickson M., Grimwood J., Schmutz J., Myers R.M., Mori S., Schluter D., Kingsley D.M. 2004. The master sex determination locus in Threespine sticklebacks is on a nascent Y chromosome. Current Biology, 14(16):1416-1424.

Penman D.J., McAndrew B.J. 2000. Genetics for the management and improvement of cultured tilapias. In: Tilapias biology and exploitation (eds. Beveridge M.C.M., McAndrew B.J.). Kluwer Academic Publishers (pp. 227-266).

Penman D.J., Piferrer F. 2008. Fish gonadogenesis. Part I: genetic and environmental mechanisms of sex determination. Reviews in Fisheries Science, 16(suppl. 1):16-34.

Penman D.J., Shah M.S., Beardmore J.A., Skibinski D.O.F. 1987. Sex ratios of gynogenetic and triploid tilapia. Proceedings of the World Symposium of Selection, Hybridization and Genetic Engineering in Aquaculture, 2:267-276.

Perez-Enriquez R., Takagi M., Taniguchi N. 1999. Genetic variability and pedigree tracing of a hatchery-reared stock of red sea bream (*Pagrus major*) used for stock enhancement, based on microsatellite DNA markers. Aquaculture, 173(1):413-423.

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.

Pfender W.F., Saha M.C., Johnson E.A., Slabaugh M.B. 2011. Mapping with RAD (restriction-site associated DNA) markers to rapidly identify QTL for stem rust resistance in *Lolium perenne*. Theoretical and Applied Genetics, 122(8):1467-1480.

Pidancier N., Miquel C., Miaud C. 2003. Buccal swabs as a non-destructive tissue sampling method for DNA analysis in amphibians. Herpetological Journal, 13(4):175-178.

Piferrer F. 2001. Endocrine sex control strategies for the feminization of teleost fish. Aquaculture, 197:229-281.

Piferrer F., Guiguen Y. 2008. Fish gonadogenesis. Part II: molecular biology and genomics of sex differentiation. Reviews in Fisheries Science, 16(suppl. 1):35-55.

Pirhonen J., Schreck C.B. 2003. Effects of anaesthesia with MS-222, clove oil and CO_2 on feed intake and plasma cortisol in Steelhead trout (*Oncorhynchus mykiss*). Aquaculture, 220:507-514.

Place A.R., Lance V.A. 2004. The TSD drama: Same cast, different stars. In: Temperature-dependent sex determination in vertebrates (eds. Valenzuela N., Lance V.). Washington, DC: Smithsonian Books (pp. 99-110).

Ponzoni R.W., Khaw H.L., Yee H.Y. 2010. GIFT: The story since leaving ICLARM (now known as The WorldFish Center). Socioeconomic, Access and Benefit Sharing and Dissemination Aspects. Fridtjof Nansen Institute. Report no. 14/2010 (p. 47).

Poonlaphdecha S., Pepey E., Canonne M., de Verdal H., Baroiller J.F., D'Cotta H. 2013. Temperature induced-masculinisation in the Nile tilapia causes rapid up-regulation of both *Dmrt1* and *Amh* expressions. General and Comparative Endocrinology, 193:234-242.

Poonlaphdecha S., Pepey E., Huang S.H., Canonne M., Soler L., Mortaji S., Morand S., Pfennig F., Mélard C., Baroiller J.F., D'Cotta H. 2011. Elevated *Amh* gene expression in the brain of male tilapia (*Oreochromis niloticus*) during testis differentiation. Sexual Development, 5(1):33-47.

Popma T.J., Ross F.E., Nerrie B.L., Bowman J.R. 1984. The development of commercial farming of tilapia in Jamaica 1979-1983. Research and Development Series No. 31. International Center for Aquaculture, Alabama Agricultural Experiment Station, Auburn University, Auburn, AL, USA (p. 18).

Pullin R.S.V., Capili J.B. 1988. Genetic improvement of tilapias: problems and prospects. The Second International Symposium on Tilapia in Aquaculture, 15:259-266.

Pullin R.S.V., Palomares M.L., Casal C.V., Dey M., Pauly D. 1997. Environmental impacts of tilapias. In: Tilapia Aquaculture, (ed. Fitzsimmons K.). Northeast Regional Agricultural Engineering Service (NRAES), Cooperative Extension, ICLARM Ithaca, NY (pp. 554-570).

Quillet E., Guillou S.L., Aubin J., Fauconneau B. 2005. Two-way selection for muscle lipid content in pan-size rainbow trout (*Onchorhynchus mykiss*). Aquaculture, 245:49-61.

Radder R.S., Quinn A.E., Georges A., Sarre S.D., Shine R. 2008. Genetic evidence for co-occurrence of chromosomal and thermal sex-determining systems in a lizard. Biology Letters, 4(2):176-178.

Raghuveer K., Senthilkumaran B. 2009. Identification of multiple *Dmrt1s* in catfish: localization, dimorphic expression pattern, changes during testicular cycle and after methyltestosterone treatment. Journal of Molecular Endocrinology, 42:437-448.

Rajaee A.H., Huntingford F.A., Ranson K.J., McAndrew B.J., Penman D.J. 2010. The effect of male colouration on reproductive success in Nile tilapia (*Oreochromis niloticus*). Aquaculture, 308:S119-S123.

Ramirez-Gonzalez R.H., Segovia V., Bird N., Fenwick P., Holdgate S., Berry S., Jack P., Caccamo M., Uauy C. 2015. RNAseq bulked segregant analysis enables the identification of high-resolution genetic markers for breeding in hexaploid wheat. Plant Biotechnology Journal, 13(5):613-624.

Ramos A.M., Crooijmans R.P.M.A., Affara N.A., Amaral A.J., Archibald A.L., Beever J.E., Bendixen C., Churcher C., Clark R., Dehais P., Hansen M.S., Hedegaard J., Hu Z.L., Kerstens H.H., Law A.S., Megens H.J., Milan D., Nonneman D.J., Rohrer G.A., Rothschild M.F., Smith T.P.L., Schnabel R.D., Tassell C.P.V., Taylor J.F., Wiedmann R.T., Schook L.B., Groenen M.A.M. 2009. Design of a high density SNP genotyping assay in the pig using SNPs identified and characterized by next generation sequencing technology. Plos One, 4(8):e6524.

Randhawa M., Bansal U., Valárik M., Klocová B., Doležel J., Bariana H. 2014. Molecular mapping of stripe rust resistance gene *Yr51* in chromosome 4AL of wheat. Theoretical and Applied Genetics, 127(2):317-324.

Reid S.M., Kidd A., Wilson C.C. 2012. Validation of buccal swabs for noninvasive DNA sampling of small-bodied imperilled fishes. Journal of Applied Ichthyology, 28:290-292.

Rengmark A.H., Slettan A., Lee W.J., Lie Ø., Lingaas F. 2007. Identification and mapping of genes associated with salt tolerance in tilapia. Journal of Fish Biology, 71(suppl. C):409-422.

Rens W., O'Brien P.C.M., Grützner F., Clarke O., Graphodatskaya D., Tsend-Ayush E., Trifonov V.A., Skelton H., Wallis M.C., Johnston S., Veyrunes F., Graves J.A.M., Ferguson-Smith M.A. 2007. The multiple sex chromosomes of Platypus and Echidna are not completely identical and several share homology with the avian Z. Genome Biology, 8(11):R243.

Rey R., Lukas-Croisier C., Lasala C., Bedecarrás P. 2003. *Amh*/mis: what we know already about the gene, the protein and its regulation. Molecular and Cellular Endocrinology, 211:21-31.

Rhoads A., Au K.F. 2015. PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics, 13(5):278-289.

Robasky K., Lewis N.E., Church G.M. 2014. The role of replicates for error mitigation in next-generation sequencing. Nature Reviews Genetics, 15:56-62.

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. Science, 326(5955):998-1001.

Rodríguez-Marí A., Yan Y.L., BreMiller R.A., Wilson C., Cañestro C., Postlethwait J.H. 2005. Characterization and expression pattern of Zebrafish anti-Müllerian hormone (*Amh*) relative to *Sox9a*, *Sox9b*, and *Cyp19a1a*, during gonad development. Gene Expression Patterns, 5(5):655-667.

Romana-Eguia M.R.R., Ikeda M., Basiao Z.U., Taniguchi N. 2005. Genetic changes during mass selection for growth in Nile tilapia, *Oreochromis niloticus* (L.), assessed by microsatellites. Aquaculture Research, 36(1):69-78.

Römer U., Beisenherz W. 1996. Environmental determination of sex in *Apistogrammai* (Cichlidae) and two other freshwater fishes (Teleostei). Journal of Fish Biology, 48(4):714-725.

Rosenstein S., Hulata G. 1994. Sex reversal in the genus *Oreochromis*: optimization of feminization protocol. Aquaculture Research, 25(3):329-339.

Rougeot C., Kanfitine S.Y., Prignon C., Gennotte V., Mélard C. 2008a. Early sex reversal during the embryonic development in the Nile tilapia. Cybium, 32:104-105.

Rougeot C., Prignon C., Kengne C.V.N., Mélard C. 2008b. Effect of high temperature during embryogenesis on the sex differentiation process in the Nile tilapia, *Oreochromis niloticus*. Aquaculture, 276:205-208.

Rye M., Gjerde B., Gjedrem T. 2010. Genetic development programs for aquaculture species in developed countries. In Proceedings of the 9th World Congress on Genetics Applied to Livestock Production, Leipzig, Germany, (p. 8).

Saccone G., Pane A. Polito L.C. 2002. Sex determination in flies, fruitflies and butterflies. Genetica, 116(1):15-23.

Saillant E., Chatain B., Menu B., Fauvel C., Vidal M.O., Fostier A. 2003. Sexual differentiation and juvenile intersexuality in the European seabass (*Dicentrarchus labrax*). Journal of Zoology, 260(1):53-63.

Sakamoto T., Danzmann R.G., Gharbi K., Howard P., Ozaki A., Khoo S.K., Woram R.A., Okamoto N., Ferguson M.M., Holm L.E., Guyomard R., Hoyheim B. 2000. A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. Genetics, 155(3):1331-1345.

Sandra G.E., Norma M.M. 2010. Sexual determination and differentiation in teleost fish. Reviews in Fish Biology and Fisheries, 20(1):101-121.

Sarder M.R.I., Penman D.J., Myers J.M., McAndrew B.J. 1999. Production and propagation of fully inbred clonal lines in the Nile tilapia (*Oreochromis niloticus* L.). Journal of Experimental Zoology, 284(6):675-685.

Sarre S.D., Georges A., Quinn A. 2004. The ends of a continuum: genetic and temperature-dependent sex determination in reptiles. BioEssays, 26(6):639-645.

Schartl M. 2004. A comparative view on sex determination in Medaka. Mechanisms of Development, 121:639-645.

Schlötterer C., Tobler R., Kofler R., Nolte V. 2014. Sequencing pools of individuals - mining genome-wide polymorphism data without big funding. Nature Reviews Genetics, 15:749-763.

Schmid M., Steinlein C. 2001. Sex chromosomes, sex-linked genes, and sex determination in the vertebrate class Amphibia. In: Genes and Mechanisms in Vertebrate Sex Determination (eds. Scherer G., Schmid M.). Birkhäuser Basel, Switzerland (pp. 143-176).

Schmid M., Steinlein C., Bogart J.P., Feichtinger W., León P., La Marca E., Díaz L.M., Sanz A., Chen S.H., Hedges S.B. 2010. The chromosomes of terraranan frogs. Insights into vertebrate cytogenetics. Cytogenetic and Genome Research, 130-131:1-568.

Schwartz M.K., Luikart G., Waples R.S. 2007. Genetic monitoring as a promising tool for conservation and management. Trends in Ecology & Evolution, 22(1):25-33.

Sekido R. 2010. *Sry*: A transcriptional activator of mammalian testis determination. The International Journal of Biochemistry & Cell Biology, 42(3):417-420.

Sekido R., Lovell-Badge R. 2009. Sex determination and *Sry*: down to a wink and a nudge? Trends in Genetics, 25(1):19-29.

Semagn K., Babu R., Hearne S., Olsen M. 2014. Single nucleotide polymorphism genotyping using kompetitive allele specific PCR (KASP): overview of the technology and its application in crop improvement. Molecular Breeding, 33(1):1-14.

Ser J.R., Roberts R.B., Kocher T.D. 2010. Multiple interacting loci control sex determination in lake Malawi cichlid fish. Evolution, 64(2):486-501.

Sham P., Bader J.S., Craig I., O'Donovan M., Owen M. 2002. DNA pooling: a tool for large-scale association studies. Nature Reviews Genetics, 3:862-871.

Shashidhar H.E., Vinod M.S., Sudhir N., Sharma G.V., Krishnamurthy K. 2005. Markers linked to grain yield using bulked segregant analysis approach in rice (*Oryza sativa* L.). Rice Genetics Newsletter, 22:69-71.

Shen M.M., Hodgkin J. 1988. *Mab-3*, a gene required for sex-specific yolk protein expression and a male-specific lineage in *C. elegans*. Cell, 54(7):1019-1031.

Shen X., Zhou M., Lu W., Ohm H. 2003. Detection of *Fusarium* head blight resistance QTL in a wheat population using bulked segregant analysis. Theoretical and Applied Genetics, 106(6):1041-1047.

Shetty S., Kirby P., Zarkower D., Graves J.A.M. 2002. *Dmrt1* in a ratite bird: evidence for a role in sex determination and discovery of a putative regulatory element. Cytogenetic and Genome Research, 99:245-251.

Shirak A., Cohen-Zinder M., Barroso R.M., Seroussi E., Ron M., Hulata G. 2009. DNA barcoding of Israeli indigenous and introduced cichlids. Israeli Journal of Aquaculture - Bamidgeh, 61(2):83-88.

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(3):1573-1581.

Shved N., Berishvili G., D'Cotta H., Baroiller J.F., Eppler E., Segner H., Reinecke M. 2007. Ethinylestradiol (EE2) differentially interferes with insulin-like growth factor I (IGF-I) in liver and extrahepatic sites during development of male and female bony fish. Journal of Endocrinology, 195:513-523.

Shved N., Berishvili G., Häusermann E., D'Cotta H., Baroiller J.F., Eppler E. 2009. Challenge with 17α-ethinylestradiol (EE2) during early development persistently impairs growth, differentiation, and local expression of IGF-I and IGF-II in immune organs of tilapia. Fish & Shellfish Immunology, 26(3):524-530.

Siegfried K.R. 2010. In search of determinants: gene expression during gonadal sex differentiation. Journal of Fish Biology, 76(8):1879-1902.

Simbolo M., Gottardi M., Corbo V., Fassan M., Mafficini A., Malpeli G., Lawlor R.T., Scarpa A. 2013. DNA qualification workflow for next generation sequencing of histopathological samples. Plos One, 8:e62692.

Simpson T.H., Johnstone R., Youngson A.F. 1976. Sex reversal in Salmonids. CM/E:48, International Council for the Exploration of the Sea, Copenhagen.

Sinclair A.H., Berta P., Palmer M.S., Hawkins J.R., Griffiths B.L., Smith M.J., Foster J.W., Frischauf A.M., Lovell-Badge R., Palmer M.S. 1990. A gene from the human sex-determining region encodes a protein with homology to a conserved DNA-binding motif. Nature, 346:240-444.

Singh A.K. 2013. Introduction of modern endocrine techniques for the production of monosex population of fishes. General and Comparative Endocrinology, 181:146-155.

Singh A.K., Pandey A.K. 1995. Effect of long photoperiod on growth, survival and sex ratio of androgenizing tilapia, *Oreochromis mossambicus*. National Academy Science Letters-India, 18(11-12):227-231.

Singh R., Singh A.K., Tripathi M. 2012. Effect of an aromatase inhibitor analogue tamoxifen on the gonad and sex differentiation in Nile tilapia *Oreochromis niloticus*. Journal of Environmental Biology, 33:799-803.

Sites Jr J.W., Bickham J.W., Haiduk M.W. 1979. Derived X chromosome in the turtle genus *Staurotypus*. Science, 206(4425):1410-1412.

Skelton P. 1993. A complete guide to the freshwater fishes of Southern Africa. Southern Book Publishers, Halfway House, Harare.

Smith C.A., Katz M., Sinclair A.H. 2003. *Dmrt1* is upregulated in the gonads during female-to-male sex reversal in ZW chicken embryos. Biology of Reproduction, 68(2):560-570.

Smith C.A., Roeszler K.N., Hudson Q.J., Sinclair A.H. 2007. Avian sex determination: what, when and where? Cytogenetic and Genome Research, 117:165-173.

Smith C.A., Roeszler K.N., Ohnesorg T., Cummins D.M., Farlie P.G., Doran T.J., Sinclair A.H. 2009. The avian Z-linked gene *Dmrt1* is required for male sex determination in the chicken. Nature, 461:267-271.

Sodsuk P., McAndrew B.J. 1991. Molecular systematics of three tilapine genera *Tilapia*, *Sarotherodon* and *Oreochromis* using allozyme data. Journal of Fish Biology, 39(sA):301-308.

Solari A.J. 1994. Sex chromosomes and sex determination in vertebrates. CRC press.

Sonesson A.K. 2007. Within-family marker-assisted selection for aquaculture species. Genetics Selection Evolution, 39(3):301-317.

Sonesson A.K., Ødegård J. 2016. Mating structures for genomic selection breeding programs in aquaculture. Genetics Selection Evolution, 48:46.

Staelens J., Rombaut D., Vercauteren I., Argue B., Benzie J., Vuylsteke M. 2008. High-density linkage maps and sex-linked markers for the black tiger shrimp (*Penaeus monodon*). Genetics, 179(2):917-925.

Steermers F.J., Gunderson K.L. 2007. Whole genome genotyping technologies on the BeadArrayTM platform. Biotechnology Journal, *2*:41-49.

Stein J., Phillips R.B., Devlin R.H. 2001. Identification of the Y chromosome in chinook salmon (*Oncorhynchus tshawytscha*). Cytogenetics and Cell Genetics, 92:108-110.

Stickney H.L., Schmutz J., Woods I.G., Holtzer C.C., Dickson M.C., Kelly P.D., Myers R.M., Talbot W.S. 2002. Rapid mapping of Zebrafish mutations with SNPs and oligonucleotide microarrays. Genome Research, 12:1929-1934.

Struussmann C.A., Moriyama S., Hanke E.F., Calsina Cota J.C., Takashima F. 1996. Evidence of thermolabile sex determination in pejerrey. Journal of Fish Biology, 48(4):643-651.

Sun Y.L., Jiang D.N., Zeng S., Hu C.J., Ye K., Yang C. Yang S.J., Li M.H., Wang D.S. 2014. Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Nile tilapia (*Oreochromis niloticus*). Aquaculture, 433(20):19-27.

Sunobe T., Nakazono A. 1993. Sex change in both directions by alteration of social dominance in *Trimma okinawae* (Pisces: Gobiidae). Ethology, 94(4):339-345.

Suzuki M.T. 2010. Sex determination: insights from the silkworm. Journal of Genetics, 89(3):357-363.

Syaifudin M. 2015. Species-specific DNA markers for improving the genetic management of tilapia. PhD Thesis, Institute of Aquaculture, University of Stirling.

Taberlet P., Waits L.P., Luikart G. 1999. Noninvasive genetic sampling: look before you leap. Trends in Ecology & Evolution, 14(8):323-327.

Takehana Y., Matsuda M., Myosho T., Suster M.L., Kawakami K., Shin-I T., Kohara Y., Kuroki Y., Toyoda A., Fujiyama A., Hamaguchi S., Sakaizumi M., Naruse K. 2014. Co-option of *Sox3* as the male-determining factor on the Y chromosome in the fish *Oryzias dancena*. Nature Communications, 5:4157.

Taniguchi N., Macaranas J.M., Pullin R.S.V. 1985. Introgressive hybridization in cultured tilapia stocks in the Philippines. Bulletin of the Japanese Society of Scientific Fisheries, 51:1219-1224.

Teixeira J., Maheswaran S., Donahoe P.K. 2001. Müllerian inhibiting substance: An instructive developmental hormone with diagnostic and possible therapeutic applications. Endocrine Reviews, 22(5):657-674.

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.

Thodesen J., Rye M., Wang Y.X., Li S.J., Bentsen H.B., Gjedrem T. 2013. Genetic improvement of tilapias in China: Genetic parameters and selection responses in growth, pond survival and cold-water tolerance of blue tilapia (*Oreochromis aureus*) after four generations of multi-trait selection. Aquaculture, 396-399:32-42.

Tonteri A., Vasemagi A., Lumme J., Primmer C.R. 2008. Use of differential expression data for identification of novel immune relevant expressed sequence tag-linked microsatellite markers in Atlantic salmon (*Salmo salar* L.). Molecular Ecology Resources, 8(6):1486-1490.

Toonen R.J., Puritz J.B., Forsman Z.H., Whitney J.L., Fernandez-Silva I., Andrews K.R., Bird C.E. 2013. ezRAD: a simplified method for genomic genotyping in non-model organisms. PeerJ, 1:e203.

Traut W., Sahara K., Marec F. 2007. Sex chromosomes and sex determination in Lepidoptera. Sexual Development, 1(6):332-346.

Trewavas E. 1983. Tilapiine Fishes of the Genera *Sarotherodon*, *Oreochromis* and *Danakilia*. British Museum (Natural History), London.

Trick M., Adamski N.M., Mugford S.G., Jiang C.C., Febrer M., Uauy C. 2012. Combining SNP discovery from next-generation sequencing data with bulked segregant analysis (BSA) to fine-map genes in polyploid wheat. BMC Plant Biology, 12:14.

Truett G.E., Heeger P., Mynatt R.L., Truett A.A., Walker J.A., Warman M.L. 2000. Preparation of PCR quality mouse genomic DNA with hot sodium hydroxide and tris (HotSHOT). BioTechniques, 29:52-54.

Turner S.D. 2014. qqman: an R package for visualizing GWAS results using QQ and manhattan plots. BiorXiv, doi: 10.1101/005165.

Turtinen L.W., Juran B.D. 1998. Protein salting-out method applied to genomic DNA isolation from fish whole blood. BioTechniques, 24:238-239.

Twongo T.K., Bayona J.D.R., Hanssens M. 2006. *Oreochromis esculentus*. The IUCN Red List of Threatened Species. www.iucnredlist.org

Uller T., Helanterä H. 2011. From the origin of sex-determining factors to the evolution of sex-determining systems. The Quarterly Review of Biology, 86(3):163-180.

Uyeno T., Miller R.R. 1971. Multiple sex chromosomes in a Mexican cyprinodontid fish. Nature, 231:452-453.

Uyeno T., Miller R.R. 1972. Second discovery of multiple sex chromosomes among fishes. Cellular and Molecular Life Sciences, 28(2):223-225.

Valenzuela N., Lance V.A. 2004. Temperature-dependent sex determination in vertebrates. Nottingham Press, Nottingham, United Kingdom (pp. 211-227).

Van der Waal B.C.W., Bills R. 2000. *Oreochromis niloticus* (Teleostei: Cichlidae) now in the Limpopo River System. South African Journal of Science, 96(1):47-48.

Van Bers N.E.M., Crooijmans R.P.M.A., Groenen M.A.M., Dibbits B.W., Komen J. 2012. SNP marker detection and genotyping in tilapia. Molecular Ecology Resources, 12(5):932-941.

Van Tassell C.P., Smith T.P.L., Matukumalli L.K., Taylor J.F., Schnabel R.D., Lawley C.T., Haudenschild C.D., Moore S.S., Warren W.C., Sonstegard T.S. 2008. SNP discovery and allele frequency estimation by deep sequencing of reduced representation libraries. Nature Methods, 5:247-252.

Vandeputte M., Dupont-Nivet M., Chavanne H., Chatain B. 2007. A polygenic hypothesis for sex determination in the European seabass *Dicentrarchus labrax*. Genetics, 176(2):1049-1057.

Varadaraj K. 1989. Feminization of *Oreochromis mossambicus* by the administration of diethylstilbestrol. Aquaculture, 80:337-341.

Veith A.M., Schäfer M., Klüver N., Schmidt C., Schultheis C., Schartl M., Winkler C., Volff J.N. 2006. Tissue-specific expression of *Dmrt* genes in embryos and adults of the platyfish *Xiphophorus maculatus*. Zebrafish, 3(3):325-337.

Verhulst E.C., van de Zande L., Beukeboom L.W. 2010. Insect sex determination: it all evolves around transformer. Current Opinion in Genetics & Development, 20(4):376-383.

Veyrunes F., Waters P.D., Miethke P., Rens W., McMillan D., Alsop A.E., Grützner F., Deakin J.E., Whittington C.M., Schatzkamer K., Kremitzki C.L., Graves T., Ferguson-Smith M.A., Warren W., Graves J.A.M. 2008. Bird-like sex chromosomes of platypus imply recent origin of mammal sex chromosomes. Genome Research, 18(6):965-973.

Vicoso B., Bachtrog D. 2013. Reversal of an ancient sex chromosome to an autosome in Drosophila. Nature, 499:332-335.

Vigier B., Forest M.G., Eychenne B., Bézard J., Garrigou O., Robel P., Josso N. 1989. Anti-Müllerian hormone produces endocrine sex reversal of fetal ovaries. Proceedings of the National Academy of Sciences of the United States of America, 86(10):3684-3688.

Villegas C.T. 1990. Growth and survival of *Oreochromis niloticus*, *O. mossambicus* and their F1 hybrids at various salinities (eds. Hirano R., Hanyu I.). The Second Asian Fisheries Forum: Proceedings of the Second Asian Fisheries Forum, Tokyo, Japan (pp. 507-510).

Vizziano D., Baron D., Randuineau G., Mahè S., Cauty C., Guiguen Y. 2008. Rainbow trout gonadal masculinization induced by inhibition of estrogen synthesis is more physiological than masculinization induced by androgen supplementation. Biology of Reproduction, 78(5):939-946.

Vizziano D., Randuineau G., Baron D., Cauty C., Guiguen Y. 2007. Characterization of early molecular sex differentiation in Rainbow trout, *Oncorhynchus mykiss*. Developmental Dynamics, 236(8):2198-2206.

Vo A.T., Jedlicka J.A. 2014. Protocols for metagenomic DNA extraction and Illumina amplicon library preparation for faecal and swab samples. Molecular Ecology Resources, 14:1183-1197.

Voelkerding K.V., Dames S.A., Durtschi J.D. 2009. Next-generation sequencing: from basic research to diagnostics. Clinical Chemistry, 55(4):641-658.

Volff J.N. 2005. Genome evolution and biodiversity in teleost fish. Heredity, 94:280-294.

Volff J.N., Nanda I., Schmid M., Schartl M. 2007. Governing sex determination in fish: regulatory putsches and ephemeral dictators. Sexual Development, 1(2):85-99.

von Hofsten J., Larsson A., Olsson P.E. 2005. Novel steroidogenic factor-1 homolog (ff1d) is coexpressed with anti-Mullerian hormone (*Amh*) in Zebrafish. Developmental Dynamics, 233(2):595-604.

Vos P., Hogers R., Bleeker M., Reijans M., van de Lee T., Hornes M., Friters A., Pot J., Paleman J., Kuiper M., Zabeau M. 1995. AFLP: a new technique for DNA fingerprinting. Nucleic Acids Research, 23(21):4407-4414.

Wang C.M., Bai Z.Y., He X.P., Lin G., Xia J.H., Sun F., Lo L.C., Feng F., Zhu Z.Y., Yue G.H. 2011. A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. BMC Genomics, 12:174.

Wang C.M., Zhu Z.Y., Lo L.C., Feng F., Lin G., Yang W.T., Li J., Yue G.H. 2007a. A microsatellite linkage map of Barramundi, *Lates calcarifer*. Genetics, 175(2):907-915.

Wang D., Mao H., Peng J., Li X., Zhou L., Gui J. 2009b. Discovery of a male-biased mutant family and identification of a male-specific SCAR marker in gynogenetic Gibel carp *Carassius auratus gibelio*. Progress in Natural Science, 19(11):1537-1544.

Wang D., Mao H.L., Chen H.X., Liu H.Q., Gui J.F. 2009a. Isolation of Y- and X-linked SCAR markers in yellow catfish and application in the production of all-male populations. Animal Genetics, 40(6):978-981.

Wang D.G., Fan J.B., Siao C.J., Berno A., Young P., Sapolsky R., Ghandour G., Perkins N., Winchester E., *et al*. 1998. Large-scale identification, mapping, and genotyping of single-nucleotide polymorphisms in the human genome. Science, 280(5366):1077-1082.

Wang D.S., Kobayashi T., Zhou L.Y., Paul-Prasanth B., Ijiri S., Sakai F., Okubo K., Morohashi K., Nagahama Y. 2007b. *Foxl2* up-regulates aromatase gene transcription in a female-specific manner by binding to the promoter as well as interacting with ad4 binding protein/steroidogenic factor 1. Molecular Endocrinology, 21(3):712-725.

Wang G.L., Paterson A.H. 1994. Assessment of DNA pooling strategies for mapping of QTLs. Theoretical and Applied Genetics, 88(3):355-361.

Wang L., Fan C., Liu Y., Zhang Y., Liu S., Sun D., Deng H., Xu Y., Tian Y., Liao X., Xie M., Li W., Chen S. 2014. A genome scan for quantitative trait loci associated with *Vibrio anguillarum* infection resistance in Japanese flounder (*Paralichthys olivaceus*) by bulked segregant analysis. Marine Biotechnology, 16(5):513-521.

Wang L.H., Tsai C.L. 2000. Effects of temperature on the deformity and sex differentiation of tilapia, *Oreochromis mossambicus*. The Journal of Experimental Zoology, 286(5):534-537.

Wang R., Sun L., Bao L., Zhang J., Jiang Y., Yao J., Song L., Feng J., Liu S., Liu Z. 2013. Bulk segregant RNA-seq reveals expression and positional candidate genes and allele-specific expression for disease resistance against enteric septicemia of catfish. BMC Genomics, 14:929.

Wasko A.P., Martins C., Oliveira C., Foresti F. 2003. Non-destructive genetic sampling in fish. An improved method for DNA extraction from fish fins and scales. Hereditas, 138:161-165.

Waters P.D., Wallis M.C., Marshall Graves J.A. 2007. Mammalian sex origin and evolution of the Y chromosome and *Sry*. Seminars in Cell & Developmental Biology, 18(3):389-400.

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.

Wessels S., Sharifi R.A., Luehmann L.M., Rueangsri S., Krause I., Pach S., Hoerstgen-Schwark G., Knorr C. 2014. Allelic variant in the anti-Müllerian hormone gene leads to autosomal and temperature-dependent sex reversal in a selected Nile tilapia line. Plos One, 9(8):e104795.

Whitefield A.K., Blaber S.J.M. 1979. The distribution of the freshwater cichlid *Sarotherodon mossambicus* in estuarine systems. Environmental Biology of Fishes, 4(1):77-81.

Wilson C.A., High S.K., McCluskey B.M., Amores A., Yan Y.L., Titus T.A., Anderson J.L., Batzel P., Carvan III M.J., Schartl M., Postlethwait J.H. 2014. Wild sex in zebrafish: loss of the natural sex determinant in domesticated strains. Genetics, 198:1291-1308.

Wilson E.B. 1905. The chromosomes in relation to the determination of sex in insects. Science, 22(564):500-502.

Wolters W.R., DeMay R. 1996. Production characteristics of Striped bass × White bass and Striped bass × Yellow bass hybrids. Journal of the World Aquaculture Society, 27(2):202-207.

Wright E.M., Snopek B., Koopman P. 1993. Seven new members of the *Sox* gene family expressed during mouse development. Nucleic Acids Research, 21(3):744.

Wright L.I., Stokes K.L., Fuller W.J., Godley B.J., McGowan A., Snape R., Tregenza T., Broderick A.C. 2012. Turtle mating patterns buffer against disruptive effects of climate change. Proceedings of the Royal Society B: Biological Sciences, 279(1736):2122-2127.

Wu G.C., Chiu P.C., Lyu Y.S., Chang C.F. 2010. The expression of *Amh* and *Amhr2* is associated with the development of gonadal tissue and sex change in the protandrous black porgy, *Acanthopagrus schlegeli*. Biology of Reproduction, 83(3):443-453.

Xia J.H., Bai Z., Meng Z., Zhang Y., Wang L., Liu F., Jing W., Wan Z.Y., Li J., Lin H., Yue G.H. 2015. Signatures of selection in tilapia revealed by whole genome resequencing. Scientific Reports, 5:14168.

Xia J.H., Liu F., Zhu Z.Y., Fu J., Feng J., Li J., Yue G.H. 2010. A consensus linkage map of the Grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs. BMC Genomics, 11:135.

Xia J.H., Wan Z.Y., Ng Z.L., Wang L., Fu G.H., Lin G., Liu F., Yue G.H. 2014. Genome-wide discovery and in silico mapping of gene-associated SNPs in Nile tilapia. Aquaculture, 432:67-73.

Yamaguchi T., Yamaguchi S., Hirai T., Kitano T. 2007. Follicle-stimulating hormone signaling and *Foxl2* are involved in transcriptional regulation of aromatase gene during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. Biochemical and Biophysical Research Communications, 359(4):935-940.

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

Yamamoto T.O. 1953. Artificially induced sex reversal in genotypic males of the Medaka (*Oryzias latipes*). Journal of Experimental Zoology, 123(3):571-594.

Yamamoto T.O. 1958. Artificial induction of functional sex-reversal in genotypic females of the Medaka (*Oryzias latipes*). Journal of Experimental Zoology, 137(2):227-263.

Yamamoto T.O. 1969. Sex differentiation. In Fish Physiology (eds. Hoar W.S., Randall D.J.). Academic Press, New York, USA (pp. 117-175).

Yang Z., Huang D., Tang W., Zheng Y., Liang K., Cutler A.J., Wu W. 2013. Mapping of quantitative trait loci underlying cold tolerance in rice seedlings via high-throughput sequencing of pooled extremes. Plos One, 8(7):e68433.

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

Yano A., Nicol B., Jouanno E., Quillet E., Fostier A., Guyomard R., Guiguen Y. 2013. The sexually dimorphic on the Y-chromosome gene (*Sdy*) is a conserved male-specific Y-chromosome sequence in many salmonids. Evolutionary Applications, 6(3):486-496.

Yoshimoto S., Ikeda N., Izutsu Y., Shiba T., Takamatsu N., Ito M. 2010. Opposite roles of *Dmrt1* and its W-linked paralogue, *Dmw*, in sexual dimorphism of *Xenopus laevis*: implications of a ZZ/ZW-type sex-determining system. Development, 137:2519-2526.

Yoshimoto S., Okada E., Umemoto H., Tamura K., Uno Y., Nishida-Umehara C., Matsuda Y., Takamatsu N., Shiba T., Ito M. 2008. A W-linked DM-domain gene, *Dmw*, participates in primary ovary development in *Xenopus laevis*. Proceedings of the National Academy of Sciences of the United States of America, 105(7):2469-2474.

Yoshinaga N., Shiraishi E., Yamamoto T., Iguchi T., Abe S., Kitano T. 2004. Sexually dimorphic expression of a teleost homologue of Müllerian inhibiting substance during gonadal sex differentiation in Japanese flounder, *Paralichthys olivaceus*. Biochemical and Biophysical Research Communications, 322(2):508-513.

Yoshiura Y., Senthilkumaran B., Watanabe M., Oba Y., Kobayashi T., Nagahama Y. 2003. Synergistic expression of Ad4BP/SF-1 and cytochrome P-450 aromatase (ovarian type) in the ovary of Nile tilapia, *Oreochromis niloticus,* during vitellogenesis suggests transcriptional interaction. Biology of Reproduction, 68(5):1545-1553.

Yue G.H. 2013. Recent advances of genome mapping and marker-assisted selection in aquaculture. Fish and Fisheries, 15(3):376-396.

Zak T., Deshev R., Benet-Perlberg A., Naor A., Magen I., Shapira Y., Ponzoni R.W., Hulata G. 2014. Genetic improvement of Israeli blue (Jordan) tilapia, *Oreochromis aureus* (Steindachner), through selective breeding for harvest weight. Aquaculture Research, 45(3):546-557.

Zhang X., Zha J., Wang Z. 2008. Influences of 4-nonylphenol on doublesex- and mab-3-related transcription factor 1 gene expression and vitellogenin mRNA induction of

adult rare minnow (*Gobiocypris rarus*). Environmental Toxicology and Chemistry, 27(1):196-205.

Zhao D., McBride D., Nandi S., McQueen H.A., McGrew M.J., Hocking P.M., Lewis P.D., Sang H.M., Clinton M. 2010. Somatic sex identity is cell autonomous in the chicken. Nature, 464:237-242.

Zhou R., Liu L., Guo Y., Yu H., Cheng H., Huang X., Tiersch T.R., Berta P. 2003. Similar gene structure of two *Sox9a* genes and their expression patterns during gonadal differentiation in a teleost fish, rice field eel (*Monopterus albus*). Molecular Reproduction and Development, 66(3):211-217.

Appendices

Chapter 3

Table C3.1 Details of the sample origin and barcode information for each sample: sample ID, sample type, source of the sample, and P1 and P2 barcode information.

Sample ID	Sample type	Fish	P1 Barcode	P2 Barcode
FWM1_1	Fin	1	TCGAG	AGTCA
SMWM1_1	Mucus	1	GTCAC	AGTCA
MWM1_1	Muscle	1	GCATT	AGTCA
FRF2_1	Fin	2	CGATA	AGTCA
SMRF2_1	Mucus	2	TGCAACA	AGTCA
MRF2_1	Muscle	2	CGTATCA	AGTCA
FWM1_2	Fin	1	ACGTA	CGATC
SMWM1_2	Mucus	1	AGAGT	CGATC
MWM1_2	Muscle	1	ATGCT	CGATC
FRF2_2	Fin	2	GACTA	CGATC
SMRF2_2	Mucus	2	CAGTCAC	CGATC
MRF2_2	Muscle	2	GCTAACA	CGATC
FWM1_3	Fin	1	TCGAG	GTCAAGT
SMWM1_3	Mucus	1	GTCAC	GTCAAGT
MWM1_3	Muscle	1	GCATT	GTCAAGT
FRF2_3	Fin	2	CGATA	GTCAAGT
SMRF2_3	Mucus	2	TGCAACA	GTCAAGT
MRF2_3	Muscle	2	CGTATCA	GTCAAGT

Chapter 4

Table C4.1 Progeny sex-ratio with the Chi-square *p*-value for the individual GIFT families used in this study.

Family	Male	Female	Total number	Sex-ratio (Male:Female)	Chi-square (*P*-value)
1	25	35	60	0.71	0.197
2	60	40	100	1.50	0.046
3	40	48	88	0.83	0.394
4	30	43	73	0.70	0.128
5	36	63	99	0.57	0.007
6	51	35	86	1.46	0.084
7	18	36	54	0.50	0.014
8	21	28	49	0.75	0.317
9	32	28	60	1.14	0.606
10	15	49	64	0.31	2.138e-05
11	21	35	56	0.60	0.061
12	22	36	58	0.61	0.067
13	31	64	95	0.48	0.0007
14	15	65	80	0.23	2.268e-08
15	22	33	55	0.67	0.138
16	17	39	56	0.44	0.003
17	23	24	47	0.96	0.884
18	23	28	51	0.82	0.484
19	40	46	86	0.87	0.518
20	1	75	76	0.01	2.2e-16
21	11	34	45	0.32	0.0006
22	59	6	65	9.83	4.903e-11
23	25	28	53	0.89	0.68
24	16	34	50	0.47	0.011
25	24	28	52	0.86	0.579
26	31	41	72	0.76	0.239
27	21	33	54	0.64	0.102
28	31	41	72	0.76	0.239

Table C4.2 Genotype data for the two Stirling families for LG1 markers (*Oni23063*, *Oni28137* and *UNH995*), and LG20 marker (*Oni3161*) with their phenotypic sex.

			LG1			**LG20**
			SNPs		**Microsatellite**	**SNP**
Sample ID	**Sex**	**Family**	***Oni23063***	***Oni28137***	***UNH995***	***Oni3161***
PF1	Dam	1	G/G	T/T	188/190	T/T
PM1	Sire	1	A/G	G/T	184/188	C/C
TK-1F	Female	1	G/G	T/T	184/190	C/C
3F	Female	1	G/G	T/T	184/188	C/C
4F	Female	1	G/G	T/T	184/190	C/C
5F	Female	1	G/G	T/T	184/190	C/C
7F	Female	1	G/G	T/T	184/190	C/C
8F	Female	1	G/G	T/T	184/188	C/T
11F	Female	1	G/G	T/T	184/188	C/C
13F	Female	1	G/G	T/T	184/188	C/C
14F	Female	1	G/G	T/T	184/188	C/T
16F	Female	1	G/G	T/T	184/190	C/C
28F	Female	1	G/G	T/T	184/188	C/C
29F	Female	1	G/G	T/T	184/190	C/T
30F	Female	1	G/G	T/T	184/188	C/C
31F	Female	1	G/G	T/T	184/188	C/C
32F	Female	1	G/G	T/T	184/190	C/C
34F	Female	1	G/G	T/T	184/190	C/T
35F	Female	1	G/G	T/T	184/190	C/T
37F	Female	1	G/G	T/T	184/190	C/C
38F	Female	1	G/G	T/T	184/190	C/T
39F	Female	1	G/G	T/T	184/188	C/T
42F	Female	1	G/G	T/T	184/190	C/T
44F	Female	1	G/G	T/T	184/190	C/T
48F	Female	1	G/G	T/T	184/188	C/C
49F	Female	1	G/G	T/T	184/188	C/T
2F	Male	1	A/G	G/T	188/190	C/C
6F	Male	1	A/G	G/T	188/190	C/T
9F	Male	1	A/G	G/T	188/190	C/C
10F	Male	1	A/G	G/T	188/190	C/C
12F	Male	1	A/G	G/T	188/188	C/C
15F	Male	1	A/G	G/T	188/188	C/C
17F	Male	1	A/G	G/T	188/188	C/C
18F	Male	1	A/G	G/T	188/190	C/T
19F	Male	1	A/G	G/T	188/190	C/T
20F	Male	1	A/G	G/T	188/188	C/C
21F	Male	1	A/G	G/T	188/190	C/C

			LG1			LG20
			SNPs		Microsatellite	SNP
Sample ID	**Sex**	**Family**	***Oni23063***	***Oni28137***	***UNH995***	***Oni3161***
22F	Male	1	A/G	G/T	188/188	C/C
23F	Male	1	A/G	G/T	188/190	C/C
24F	Male	1	A/G	G/T	188/188	C/T
25F	Male	1	A/G	G/T	188/188	C/T
26F	Male	1	A/G	G/T	188/188	C/T
27F	Male	1	A/G	G/T	188/188	C/T
33F	Male	1	A/G	G/T	188/190	C/T
36F	Male	1	A/G	G/T	188/190	C/T
40F	Male	1	A/G	G/T	188/190	C/C
41F	Male	1	A/G	G/T	188/190	C/C
43F	Male	1	A/G	G/T	188/188	C/T
45F	Male	1	A/G	G/T	188/188	C/C
46F	Male	1	A/G	G/T	188/188	C/T
47F	Male	1	A/G	G/T	188/188	C/T
50F	Male	1	A/G	G/T	188/188	C/T
DOFCF	Dam	2	G/G	T/T	184/188	C/T
CE248	Sire	2	A/G	G/T	188/192	C/T
1CB1	Female	2	G/G	T/T	184/192	C/C
2CB1	Female	2	G/G	T/T	188/192	T/T
4CB1	Female	2	G/G	T/T	188/192	C/T
6CB1	Female	2	G/G	T/T	188/192	C/T
8CB1	Female	2	G/G	T/T	184/192	T/T
9CB1	Female	2	G/G	T/T	188/192	C/T
11CB1	Female	2	G/G	T/T	184/192	C/C
14CB1	Female	2	G/G	T/T	188/192	C/T
18CB1	Female	2	G/G	T/T	188/192	T/T
19CB1	Female	2	G/G	T/T	188/192	C/T
20CB1	Female	2	G/G	T/T	184/192	C/T
21CB1	Female	2	G/G	T/T	188/192	C/T
23CB1	Female	2	G/G	T/T	184/188	C/T
24CB1	Female	2	G/G	T/T	188/192	C/C
31CB1	Female	2	G/G	T/T	188/192	C/T
34CB1	Female	2	G/G	T/T	184/192	C/C
36CB1	Female	2	G/G	T/T	188/192	C/T
39CB1	Female	2	G/G	T/T	184/192	C/T
40CB1	Female	2	G/G	T/T	184/192	C/C
41CB1	Female	2	G/G	T/T	184/192	C/T
43CB1	Female	2	G/G	T/T	184/192	C/T
44CB1	Female	2	G/G	T/T	184/192	C/T
45CB1	Female	2	G/G	T/T	184/192	C/T

			LG1			**LG20**
			SNPs		**Microsatellite**	**SNP**
Sample ID	**Sex**	**Family**	***Oni23063***	***Oni28137***	***UNH995***	***Oni3161***
48CB1	Female	2	G/G	T/T	188/192	C/T
51CB1	Female	2	G/G	T/T	188/192	T/T
54CB1	Female	2	G/G	T/T	184/192	C/T
55CB1	Female	2	G/G	T/T	188/192	T/T
56CB1	Female	2	G/G	T/T	184/192	C/T
57CB1	Female	2	G/G	T/T	184/192	C/T
3CB1	Male	2	A/G	G/T	184/188	C/T
5CB1	Male	2	A/G	G/T	188/188	C/C
7CB1	Male	2	A/G	G/T	184/188	C/C
10CB1	Male	2	A/G	G/T	184/188	C/T
12CB1	Male	2	A/G	G/T	184/188	T/T
13CB1	Male	2	A/G	G/T	184/188	T/T
15CB1	Male	2	A/G	G/T	184/188	C/C
16CB1	Male	2	A/G	G/T	184/188	T/T
17CB1	Male	2	A/G	G/T	188/188	C/T
22CB1	Male	2	A/G	G/T	188/188	C/C
25CB1	Male	2	A/G	G/T	188/188	C/T
26CB1	Male	2	A/G	G/T	188/188	C/T
27CB1	Male	2	A/G	G/T	188/188	C/C
28CB1	Male	2	A/G	G/T	188/188	C/T
29CB1	Male	2	A/G	G/T	188/188	C/C
30CB1	Male	2	A/G	G/T	188/188	T/T
32CB1	Male	2	A/G	G/T	188/188	C/T
33CB1	Male	2	A/G	G/T	184/188	C/T
35CB1	Male	2	A/G	G/T	184/188	C/T
37CB1	Male	2	A/G	G/T	188/188	C/T
38CB1	Male	2	A/G	G/T	188/188	C/C
42CB1	Male	2	A/G	G/T	188/188	C/T
46CB1	Male	2	A/G	G/T	184/192	T/T
47CB1	Male	2	A/G	G/T	188/188	T/T
49CB1	Male	2	A/G	G/T	188/188	C/T
50CB1	Male	2	A/G	G/T	184/188	C/C
52CB1	Male	2	A/G	G/T	188/188	C/C
53CB1	Male	2	A/G	G/T	188/188	C/C
58CB1	Male	2	A/G	G/T	188/188	C/T

Table C4.3 Details of each sample used for BSA-ddRADseq: sample ID, sex, source of the sample, ddRADseq run, P1 and P2 barcode information and the generated paired-end reads.

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
09Dam_1	Dam	GIFT	1	1	TCAGA	TAGCA	123643
09Sire_1	Sire	GIFT	1	1	GATCG	TAGCA	157298
09FPP1_1	Female progeny pool	GIFT	1	1	CATGA	TAGCA	95965
09MPP1_1	Male progeny pool	GIFT	1	1	ATCGA	TAGCA	142525
12Dam_1	Dam	GIFT	2	1	TCGAG	TAGCA	126529
12Sire_1	Sire	GIFT	2	1	GTCAC	TAGCA	121730
12FPP1_1	Female progeny pool	GIFT	2	1	GCATT	TAGCA	163289
12MPP1_1	Male progeny pool	GIFT	2	1	CGATA	TAGCA	222765
13Dam_1	Dam	GIFT	3	1	TGCAACA	TAGCA	142380
13Sire_1	Sire	GIFT	3	1	CGTATCA	TAGCA	190345
13FPP2_1	Female progeny pool	GIFT	3	1	CACAGAC	TAGCA	158195
13MPP2_1	Male progeny pool	GIFT	3	1	ACTGCAC	TAGCA	134364
15Dam_1	Dam	GIFT	4	1	TCTCTCA	AGCTGTC	111468
15Sire_1	Sire	GIFT	4	1	GTACACA	AGCTGTC	137183
15FPP2_1	Female progeny pool	GIFT	4	1	CTCTTCA	AGCTGTC	157562
15MPP2_1	Male progeny pool	GIFT	4	1	CTAGGAC	AGCTGTC	150490
23Dam_1	Dam	GIFT	5	1	ACGTA	AGCTGTC	189657
23Sire_1	Sire	GIFT	5	1	AGAGT	AGCTGTC	160752
23FPP1_1	Female progeny pool	GIFT	5	1	ATGCT	AGCTGTC	126366
23MPP1_1	Male progeny pool	GIFT	5	1	GACTA	AGCTGTC	165290
C1PF_1	Dam	Stirling	1	1	CAGTCAC	AGCTGTC	98493
C1PM_1	Sire	Stirling	1	1	GCTAACA	AGCTGTC	188991

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
C1FPP1_1	Female progeny pool	Stirling	1	1	ACACGAG	AGCTGTC	162278
C1MPP1_1	Male progeny pool	Stirling	1	1	AGGACAC	AGCTGTC	144365
C2PF_1	Dam	Stirling	2	1	TCAGA	AGTCA	16273
C2PM_1	Sire	Stirling	2	1	GATCG	AGTCA	82445
C2FPP1_1	Female progeny pool	Stirling	2	1	CATGA	AGTCA	90763
C2MPP1_1	Male progeny pool	Stirling	2	1	ATCGA	AGTCA	111798
09Dam_2	Dam	GIFT	1	1	TCTCTCA	TACGTGT	81705
9Sire_2	Sire	GIFT	1	1	GTACACA	TACGTGT	86160
9FPP1_2	Female progeny pool	GIFT	1	1	CTCTTCA	TACGTGT	61668
9MPP1_2	Male progeny pool	GIFT	1	1	CTAGGAC	TACGTGT	78827
12Dam_2	Dam	GIFT	2	1	ACGTA	TACGTGT	113777
12Sire_2	Sire	GIFT	2	1	AGAGT	TACGTGT	203213
12FPP1_2	Female progeny pool	GIFT	2	1	ATGCT	TACGTGT	84532
12MPP1_2	Male progeny pool	GIFT	2	1	GACTA	TACGTGT	83973
13Dam_2	Dam	GIFT	3	1	CAGTCAC	TACGTGT	99497
13Sire_2	Sire	GIFT	3	1	GCTAACA	TACGTGT	131533
13FPP2_2	Female progeny pool	GIFT	3	1	ACACGAG	TACGTGT	125036
13MPP2_2	Male progeny pool	GIFT	3	1	AGGACAC	TACGTGT	71009
15Dam_2	Dam	GIFT	4	1	TCAGA	GCATA	84256
15Sire_2	Sire	GIFT	4	1	GATCG	GCATA	81829
15FPP2_2	Female progeny pool	GIFT	4	1	CATGA	GCATA	106166
15MPP2_2	Male progeny pool	GIFT	4	1	ATCGA	GCATA	139187
23Dam_2	Dam	GIFT	5	1	TCGAG	GCATA	100642

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
23Sire_2	Sire	GIFT	5	1	GTCAC	GCATA	87363
23FPP1_2	Female progeny pool	GIFT	5	1	GCATT	GCATA	91943
23MPP1_2	Male progeny pool	GIFT	5	1	CGATA	GCATA	128612
C1PF_2	Dam	Stirling	1	1	TGCAACA	GCATA	84586
C1PM_2	Sire	Stirling	1	1	CGTATCA	GCATA	117617
C1FPP1_2	Female progeny pool	Stirling	1	1	CACAGAC	GCATA	118846
C1MPP1_2	Male progeny pool	Stirling	1	1	ACTGCAC	GCATA	113788
C2PF_2	Dam	Stirling	2	1	TCTCTCA	CGATC	15307
C2PM_2	Sire	Stirling	2	1	GTACACA	CGATC	69706
C2FPP1_2	Female progeny pool	Stirling	2	1	CTCTTCA	CGATC	120063
C2MPP1_2	Male progeny pool	Stirling	2	1	CTAGGAC	CGATC	98593
09Dam_3	Dam	GIFT	1	1	TCAGA	CATCTGT	157346
9Sire_3	Sire	GIFT	1	1	GATCG	CATCTGT	173931
9FPP1_3	Female progeny pool	GIFT	1	1	CATGA	CATCTGT	139715
9MPP1_3	Male progeny pool	GIFT	1	1	ATCGA	CATCTGT	158759
12Dam_3	Dam	GIFT	2	1	TCGAG	CATCTGT	202028
12Sire_3	Sire	GIFT	2	1	GTCAC	CATCTGT	176908
12FPP1_3	Female progeny pool	GIFT	2	1	GCATT	CATCTGT	200232
12MPP1_3	Male progeny pool	GIFT	2	1	CGATA	CATCTGT	242729
13Dam_3	Dam	GIFT	3	1	TGCAACA	CATCTGT	175317
13Sire_3	Sire	GIFT	3	1	CGTATCA	CATCTGT	233969
13FPP2_3	Female progeny pool	GIFT	3	1	CACAGAC	CATCTGT	197617
13MPP2_3	Male progeny pool	GIFT	3	1	ACTGCAC	CATCTGT	177787
15Dam_3	Dam	GIFT	4	1	TCTCTCA	CTGGT	107502

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
15Sire_3	Sire	GIFT	4	1	GTACACA	CTGGT	109129
15FPP2_3	Female progeny pool	GIFT	4	1	CTCTTCA	CTGGT	130741
15MPP2_3	Male progeny pool	GIFT	4	1	CTAGGAC	CTGGT	154604
23Dam_3	Dam	GIFT	5	1	ACGTA	CTGGT	137201
23Sire_3	Sire	GIFT	5	1	AGAGT	CTGGT	119525
23FPP1_3	Female progeny pool	GIFT	5	1	ATGCT	CTGGT	121042
23MPP1_3	Male progeny pool	GIFT	5	1	GACTA	CTGGT	115965
C1PF_3	Dam	Stirling	1	1	CAGTCAC	CTGGT	89216
C1PM_3	Sire	Stirling	1	1	GCTAACA	CTGGT	143662
C1FPP1_3	Female progeny pool	Stirling	1	1	ACACGAG	CTGGT	121026
C1MPP1_3	Male progeny pool	Stirling	1	1	AGGACAC	CTGGT	85784
C2PF_3	Dam	Stirling	2	1	TCAGA	GTCAAGT	43332
C2PM_3	Sire	Stirling	2	1	GATCG	GTCAAGT	118512
C2FPP1_3	Female progeny pool	Stirling	2	1	CATGA	GTCAAGT	160487
C2MPP1_3	Male progeny pool	Stirling	2	1	ATCGA	GTCAAGT	207241
09FPP1_4	Female progeny pool	GIFT	1	1	TCTCTCA	GAAGC	85791
09MPP1_4	Male progeny pool	GIFT	1	1	GTACACA	GAAGC	86258
12FPP1_4	Female progeny pool	GIFT	2	1	CTCTTCA	GAAGC	88832
12MPP1_4	Male progeny pool	GIFT	2	1	CTAGGAC	GAAGC	139424
13FPP2_4	Female progeny pool	GIFT	3	1	ACGTA	GAAGC	169902
13MPP2_4	Male progeny pool	GIFT	3	1	AGAGT	GAAGC	171811
15FPP2_4	Female progeny pool	GIFT	4	1	ATGCT	GAAGC	138987
15MPP2_4	Male progeny pool	GIFT	4	1	GACTA	GAAGC	122574
23FPP1_4	Female progeny pool	GIFT	5	1	CAGTCAC	GAAGC	146200

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
23MPP1_4	Male progeny pool	GIFT	5	1	GCTAACA	GAAGC	147257
C1FPP1_4	Female progeny pool	Stirling	1	1	ACACGAG	GAAGC	203344
C1MPP1_4	Male progeny pool	Stirling	1	1	AGGACAC	GAAGC	159651
C2FPP1_4	Female progeny pool	Stirling	2	1	CACAGAC	ATACGGT	161103
C2MPP1_4	Male progeny pool	Stirling	2	1	ACTGCAC	ATACGGT	144177
025Dam	Dam	GIFT	6	2	TCAGA	GCATA	112012
025Sire	Sire	GIFT	6	2	TGCAACA	GAGATGT	153459
025FPP	Female progeny pool	GIFT	6	2	GATCG	CGATC	241839
025MPP	Male progeny pool	GIFT	6	2	CGTATCA	CATCTGT	136387
027Dam	Dam	GIFT	7	2	CATGA	CTGGT	143505
027Sire	Sire	GIFT	7	2	CACAGAC	GTCAAGT	188324
027FPP	Female progeny pool	GIFT	7	2	ATCGA	GAAGC	204086
027MPP	Male progeny pool	GIFT	7	2	ACTGCAC	ATACGGT	219802
031Dam	Dam	GIFT	8	2	TCGAG	TAGCA	180441
031Sire	Sire	GIFT	8	2	TCTCTCA	AGCTGTC	147983
031FPP	Female progeny pool	GIFT	8	2	GTCAC	AGTCA	155088
031MPP	Male progeny pool	GIFT	8	2	GTACACA	TACGTGT	109642
033Dam	Dam	GIFT	9	2	GCATT	GCATA	161261
033FPP	Female progeny pool	GIFT	9	2	CTCTTCA	GAGATGT	136975
033MPP	Male progeny pool	GIFT	9	2	CGATA	CGATC	229746
037Dam	Dam	GIFT	10	2	CTAGGAC	CATCTGT	248577
037Sire	Sire	GIFT	10	2	ACGTA	CTGGT	278181
037FPP	Female progeny pool	GIFT	10	2	CAGTCAC	GTCAAGT	162717
037MPP	Male progeny pool	GIFT	10	2	AGAGT	GAAGC	186951

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
044Dam	Dam	GIFT	11	2	GCTAACA	ATACGGT	144088
044Sire	Sire	GIFT	11	2	ATGCT	TAGCA	114381
044FPP	Female progeny pool	GIFT	11	2	ACACGAG	AGCTGTC	221514
044MPP	Male progeny pool	GIFT	11	2	GACTA	AGTCA	208800
050Dam	Dam	GIFT	12	2	AGGACAC	TACGTGT	192117
050Sire	Sire	GIFT	12	2	TCAGA	GAGATGT	228102
050FPP	Female progeny pool	GIFT	12	2	TGCAACA	CGATC	140324
050MPP	Male progeny pool	GIFT	12	2	GATCG	CATCTGT	174325
052Dam	Dam	GIFT	13	2	CGTATCA	CTGGT	188587
052Sire	Sire	GIFT	13	2	CATGA	GTCAAGT	245158
052FPP	Female progeny pool	GIFT	13	2	CACAGAC	GAAGC	118119
052MPP	Male progeny pool	GIFT	13	2	ATCGA	ATACGGT	181093
053Dam	Dam	GIFT	14	2	ACTGCAC	TAGCA	179638
053Sire	Sire	GIFT	14	2	TCGAG	AGCTGTC	194859
053FPP	Female progeny pool	GIFT	14	2	TCTCTCA	AGTCA	196819
053MPP	Male progeny pool	GIFT	14	2	GTCAC	TACGTGT	141913
056Dam	Dam	GIFT	15	2	GTACACA	GCATA	176234
056Sire	Sire	GIFT	15	2	GCATT	GAGATGT	203256
056FPP	Female progeny pool	GIFT	15	2	CTCTTCA	CGATC	181425
056MPP	Male progeny pool	GIFT	15	2	CGATA	CATCTGT	172846
087Dam	Dam	GIFT	16	2	CTAGGAC	CTGGT	228579
087Sire	Sire	GIFT	16	2	ACGTA	GTCAAGT	207664
087FPP	Female progeny pool	GIFT	16	2	CAGTCAC	GAAGC	211260
087MPP	Male progeny pool	GIFT	16	2	AGAGT	ATACGGT	138363

Sample ID	Gender	Source	Family	Run	P1 Barcode	P2 Barcode	Paired-end reads
097Dam1	Dam	GIFT	17	2	GCTAACA	TAGCA	301583
097Dam2	Dam	GIFT	17	2	ATGCT	AGTCA	192209
097Sire1	Sire	GIFT	17	2	ATGCT	AGCTGTC	183624
097Sire2	Sire	GIFT	17	2	ACACGAG	TACGTGT	177809
097FPP	Female progeny pool	GIFT	17	2	ACACGAG	AGTCA	172217
097MPP	Male progeny pool	GIFT	17	2	GACTA	TACGTGT	168801
101Dam	Dam	GIFT	18	2	AGGACAC	GCATA	141976
101Sire	Sire	GIFT	18	2	TCAGA	CGATC	178195
101FPP	Female progeny pool	GIFT	18	2	TGCAACA	CATCTGT	176494
101MPP	Male progeny pool	GIFT	18	2	GATCG	CTGGT	154099
105Dam	Dam	GIFT	19	2	CGTATCA	GTCAAGT	173236
105Sire	Sire	GIFT	19	2	CATGA	GAAGC	259019
105FPP	Female progeny pool	GIFT	19	2	CACAGAC	ATACGGT	204079
105MPP	Male progeny pool	GIFT	19	2	ATCGA	TAGCA	292005

Table C4.4 KASP assay sequences used in Chapter 3, 4, 5 and 6.

Assay ID	Primer_AlleleFAM	Primer_AlleleHEX	Primer_Common	Allele FAM	Allele HEX
ss2017360173	AAACACCAAATATACC TTAAACTGTGATCA	CACCAAATATACCTT AAACTGTGATCG	TGTAAACTCTGAAGTG ATATTAAGGTGTTT	T	C
ss2017360175	GCTCTGGAGTATCTGC TGTA	GCTGCTCTGGAGTAT CTGCTGTT	ACACTGAACCCAGCAG ACCAGAATA	A	T
ss2017360134	TTACACTAGAAATCAA AGGTTAATGACAG	TACACTAGAAATCAA AGGTTAATGACAA	CATTCAGTTTAGACTCA GAAATCCACATTT	C	T
ss2017360178	TAAAAGAAAATCTACC GATACTGAATTCATA	AAAAGAAAATCTACC GATACTGAATTCATG	GATTGCTAGTTYTGTGA CACAGATTCATTT	A	G
ss2017360168	GTTGTTTCGTTATGATA AAGATGGGG	TTGTTGTTTCGTTATG ATAAAGATGGGA	CTGCAGGAGTCAAACT GTGCAATAATTAA	C	T
ss831884014	AACTGAGTGCGTTACA GGAGAAAG	AACTGAGTGCGTTAC AGGAGAAAC	TTGCACATGTCACCTGT GGCATGTT	G	C
Oni28137	ACCAAGACGCCACAGA CAGTTG	CACCAAGACGCCACA GACAGTTT	GAGACCGTGGCGTCAG ACAGTA	G	T
Oni23063	AAAGTGAAATCCCAGC CACA	GCTAAAGTGAAATCC CAGCCACG	TGCTGAACGCRTCCTCA AACATTACAT	A	G
Oni3161	GAAGGTGACCAAGTTC ATGCTGTTACTTTTCTC TTTGAGTTATTTTAGTT AGC	GAAGGTCGGAGTCAA CGGATTAGTTACTTTT CTCTTTGAGTTATTTT AGTTAGT	GCCCCAGCAATTATAA AATTACCACTTAAA	C	T

Table C4.5 Primer sequences used in Chapter 3, 4, 5, 6.

Marker ID	Forward sequences	Reverse sequences
UNH995	CCAGCCCTCTGCATAAAGAC	GCAGCACAACCACAGTGCTA
UNH898	GATGTCCCCACAAGGTATGAA	TAATCCACTCACCCCGTTTC
ARO172	AGGCCTTTCATCGCTGTTTT	ACCCTGTAGATGAGCGCAAA
Amh exon VII	AGCAGCTCTAGCGGCATCCACA	TGTGTTTTCTTTCTGCGTCCGCCA
Amh exon VI	AAACCTCCTTCCTTTGTGAATGTC	CGTGGCCACTCCCTCCACCC
*Amh*_E0	GAGCTGGGTTGGAGTCATTG	ACAGGTCCCGGAAAGAAAGTT
*Amh*_E0_del	AGTGAGCTGGGTTGGAGTCATT	CGTCGACAACAAAGCTCAAAAC
*Amhy*_E0_E2	GGACATCCCCCGCTTAGAGAA	ATGCGGTTTCAGCTTTTACCTG
*Amhy*_Promoter_del	GAAAGGGGTGTTTTGGTGCTGGC	ACCCAGGAAGCGTTTCATCTCA
*Amh*_SNP_exon_VI	GAGGTTTCACTGGGAGCCAA	TACTTACATGCACCCGACCG

Table C4.6 Phenotypic sex and genotype information for each marker studied for six GIFT families, GIFT and Stirling broodstock.

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
1	**1D**	**Dam**	**X/Y**	**T/T**	**A/A**	**C/C**	**G/G**	**C/C**	**267/285**	**252/274**	**253/256**
1	**1S**	**Sire**	**X/X**	**C/C**	**A/A**	**C/C**	**A/A**	**C/T**	**259/301**	**246/340**	**253/253**
1	3F09	Female	X/X					C/C	285/301	252/340	
1	8F09	Female	X/X					C/C	285/301	252/340	
1	10F09	Female	X/X					C/C	285/301	252/340	
1	12F09	Female	X/X					C/C	285/301	252/340	
1	13F09	Female	X/X					C/T	259/285	246/252	
1	14F09	Female	X/X					C/C	285/301	252/340	
1	15F09	Female	X/Y					C/C	267/301	274/340	
1	17F09	Female	X/X					C/C	285/301	252/340	
1	21F09	Female	X/X					C/T	259/285	246/252	
1	23F09	Female	X/X					C/T	259/285	246/252	
1	24F09	Female	X/X					C/T	259/285	246/252	
1	26F09	Female	X/X					C/C	285/301	252/340	
1	28F09	Female	X/X					C/C	285/301	252/340	
1	32F09	Female	X/X					C/T	259/285	246/252	
1	33F09	Female	X/X					C/T	259/285	246/252	
1	38F09	Female	X/X					C/T	259/285	246/252	
1	40F09	Female	X/X					C/C	285/301	252/340	
1	42F09	Female	X/X					C/T	259/285	246/252	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
1	43F09	Female	X/X					C/C	285/301	252/340	
1	50F09	Female	X/X					C/T	259/285	246/252	
1	2M09	Male	X/Y					C/T	259/267	246/274	
1	7M09	Male	X/Y					C/C	267/301	274/340	
1	9M09	Male	X/Y					C/T	259/267	246/274	
1	11M09	Male	X/Y					C/C	267/301	274/340	
1	16M09	Male	X/X					C/C	285/301	252/340	
1	18M09	Male	X/Y					C/T	259/267	246/274	
1	20M09	Male	X/Y					C/T	259/267	246/274	
1	25M09	Male	X/Y					C/T	259/267	246/274	
1	27M09	Male	X/Y					C/T	259/267	246/274	
1	29M09	Male	X/Y					C/C	267/301	274/340	
1	31M09	Male	X/X					C/T	259/285	246/252	
1	34M09	Male	X/Y					C/T	259/267	246/274	
1	35M09	Male	X/X					C/C	285/301	252/340	
1	41M09	Male	X/Y					C/C	267/301	274/340	
1	44M09	Male	X/Y					C/C	267/301	274/340	
1	46M09	Male	X/X					C/C	285/301	252/340	
1	47M09	Male	X/Y					C/C	267/301	274/340	
1	51M09	Male	X/Y					C/T	259/267	246/274	
1	54M09	Male	X/X					C/T	259/285	246/252	
1	57M09	Male	X/Y					C/C	267/301	274/340	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
2	**2D**	**Dam**	**X/X**	**C/C**	**T/T**	**C/C**	**G/G**	**C/C**	**269/269**	**276/276**	**253/253**
2	**2S**	**Sire**	**X/Y**	**C/T**	**A/A**	**C/T**	**G/G**	**C/T**	**259/267**	**246/274**	**253/256**
2	12F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	16F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	17F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	19F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	29F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	32F12	Female	X/X	C/T		C/C		C/T	259/269	246/276	
2	35F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	37F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	38F12	Female	X/Y	C/T		C/T		C/C	267/269	274/276	
2	45F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	48F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	49F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	50F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	51F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	52F12	Female	X/X	C/C		C/T		C/T	259/269	246/276	
2	53F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	58F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	65F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	67F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	72F12	Female	X/X	C/C		C/C		C/T	259/269	246/276	
2	1M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
2	2M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	3M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
2	6M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	7M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	8M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
2	9M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	14M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
2	15M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	18M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	20M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	21M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	22M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	23M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	24M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	25M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	33M12	Male	X/Y	C/T		C/T		C/C	267/269	274/276	
2	41M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
2	42M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
2	43M12	Male	X/X	C/C		C/C		C/T	259/269	246/276	
3	**3D**	**Dam**	**X/X**	**C/C**	**A/A**	**C/C**	**A/G**	**C/C**	**275/301**	**244/336**	**253/253**
3	**3S**	**Sire**	**X/Y**	**C/T**	**A/A**	**C/T**	**G/G**	**C/T**	**259/267**	**246/274**	**253/256**
3	4F13	Female	X/X	C/C		C/C	G/G	C/T	259/301	246/336	
3	5F13	Female	X/X	C/C		C/C	G/G	C/T	259/275	246/336	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
3	7F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	9F13	Female	X/X	C/C		C/C	A/G	C/T	259/301	244/246	
3	13F13	Female	X/X	C/C		C/C	G/G	C/T	259/275	246/336	
3	36F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	42F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	44F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	46F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	47F13	Female	X/X	C/C		C/C	G/G	C/T	259/301	246/336	
3	49F13	Female	X/X	C/C		C/C	G/G	C/T	259/301	246/336	
3	50F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	51F13	Female	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	52F13	Female	X/X	C/C		C/C	G/G	C/T	259/301	246/336	
3	55F13	Female	X/X	C/C		C/C	G/G	C/T	259/301	246/336	
3	66F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	71F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	72F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	74F13	Female	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	85F13	Female	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	2M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	3M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	12M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	15M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	18M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
3	27M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	29M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	30M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	33M13	Male	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	37M13	Male	X/Y	C/T		C/C	G/G	C/C	267/301	274/336	
3	38M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	40M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	41M13	Male	X/Y	C/T		C/T	A/G	C/C	267/301	274/336	
3	43M13	Male	X/Y	C/C		C/T	G/G	C/C	267/301	274/336	
3	45M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	48M13	Male	X/Y	C/T		C/C	A/G	C/C	267/275	244/274	
3	53M13	Male	X/Y	C/T		C/T	G/G	C/C	267/301	274/336	
3	56M13	Male	X/X	C/C		C/C	A/G	C/T	259/275	244/246	
3	57M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
3	65M13	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
4	**4D**	**Dam**	**X/X**	**C/T**	**A/T**	**C/C**	**A/G**	**C/C**	**275/301**	**244/336**	**253/253**
4	**4S**	**Sire**	**X/Y**	**C/T**	**A/A**	**C/C**	**A/A**	**C/C**	**267/275**	**244/274**	**253/256**
4	9F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	20F15	Female	X/X	C/T	A/A		A/A		275/301	244/244	
4	25F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	26F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	27F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	30F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
4	31F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	32F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	35F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	39F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	40F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	41F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	42F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	43F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	46F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	47F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	50F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	51F15	Female	X/X	C/T	A/A		A/A		275/301	244/336	
4	58F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	67F15	Female	X/X	C/C	A/T		A/G		275/275	244/244	
4	2M15	Male	X/Y	C/T	A/A		A/A		267/275	244/274	
4	3M15	Male	X/Y	C/T	A/T		A/G		267/301	274/336	
4	6M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	10M15	Male	X/Y	T/T	A/A		A/A		267/301	274/336	
4	15M15	Male	X/Y	T/T	A/A		A/A		267/275	244/274	
4	18M15	Male	X/Y	C/T	A/T		A/G		267/301	244/274	
4	19M15	Male	X/Y	T/T	A/A		A/A		267/275	274/336	
4	21M15	Male	X/Y	C/T	A/T		A/G		267/301	244/274	
4	22M15	Male	X/Y	T/T	A/A		A/A		267/275	274/336	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
4	33M15	Male	X/Y	C/T	A/T		A/G		267/301	244/274	
4	34M15	Male	X/Y	T/T	A/A		A/A		267/275	274/336	
4	37M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	52M15	Male	X/Y	T/T	A/A		A/A		267/301	274/336	
4	53M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	54M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	55M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	56M15	Male	X/Y	T/T	A/A		A/A		267/301	274/336	
4	65M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	71M15	Male	X/Y	C/T	A/T		A/G		267/275	244/274	
4	72M15	Male	X/Y	T/T	A/A		A/A		267/301	274/336	
7	**7D**	**Dam**	**X/X**	**C/C**	**A/A**	**C/C**	**A/G**	**C/C**	**275/301**	**244/336**	**253/253**
7	**7S**	**Sire**	**X/Y**	**C/T**	**A/T**	**C/C**	**G/G**	**C/C**	**267/269**	**274/276**	**253/256**
7	2F27	Female	X/X	C/C	A/T		G/G		269/301	276/336	
7	4F27	Female	X/X	C/C	A/T		G/G		269/301	276/336	
7	6F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	7F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	12F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	14F27	Female	X/X	C/C	A/T		A/G		269/301	244/276	
7	16F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	19F27	Female	X/X	C/C	A/T		A/G		269/301	276/336	
7	22F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	23F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
7	24F27	Female	X/X	C/C	A/T		G/G		269/301	276/336	
7	25F27	Female	X/X	C/C	A/T		G/G		269/301	276/336	
7	26F27	Female	X/X	C/C	A/T		G/G		269/301	276/336	
7	27F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	28F27	Female	X/X	C/C	A/T		A/G		269/275	244/276	
7	1M27	Male	X/Y	C/T	A/A		A/G		267/301	244/274	
7	3M27	Male	X/Y	C/T	A/A		G/G		267/275	274/336	
7	5M27	Male	X/Y	C/T	A/A		A/G		267/301	244/274	
7	8M27	Male	X/Y	C/T	A/A		G/G		267/275	274/336	
7	9M27	Male	X/Y	C/T	A/A		A/G		267/301	244/274	
7	10M27	Male	X/Y	C/T	A/A		A/G		267/275	274/336	
7	11M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	13M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	20M27	Male	X/Y	C/T	A/A		G/G		267/301	274/336	
7	29M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	33M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	37M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	41M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
7	52M27	Male	X/Y	C/C	A/T		A/G		267/275	244/274	
7	53M27	Male	X/Y	C/T	A/A		A/G		267/275	244/274	
19	**19D**	**Dam**	**X/X**	**C/C**	**A/A**	**C/C**	**G/G**	**C/C**	**269/275**	**244/276**	**253/253**
19	**19S**	**Sire**	**X/Y**	**C/T**	**A/A**	**C/T**	**A/G**	**C/T**	**259/267**	**246/274**	**253/256**
19	1F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	

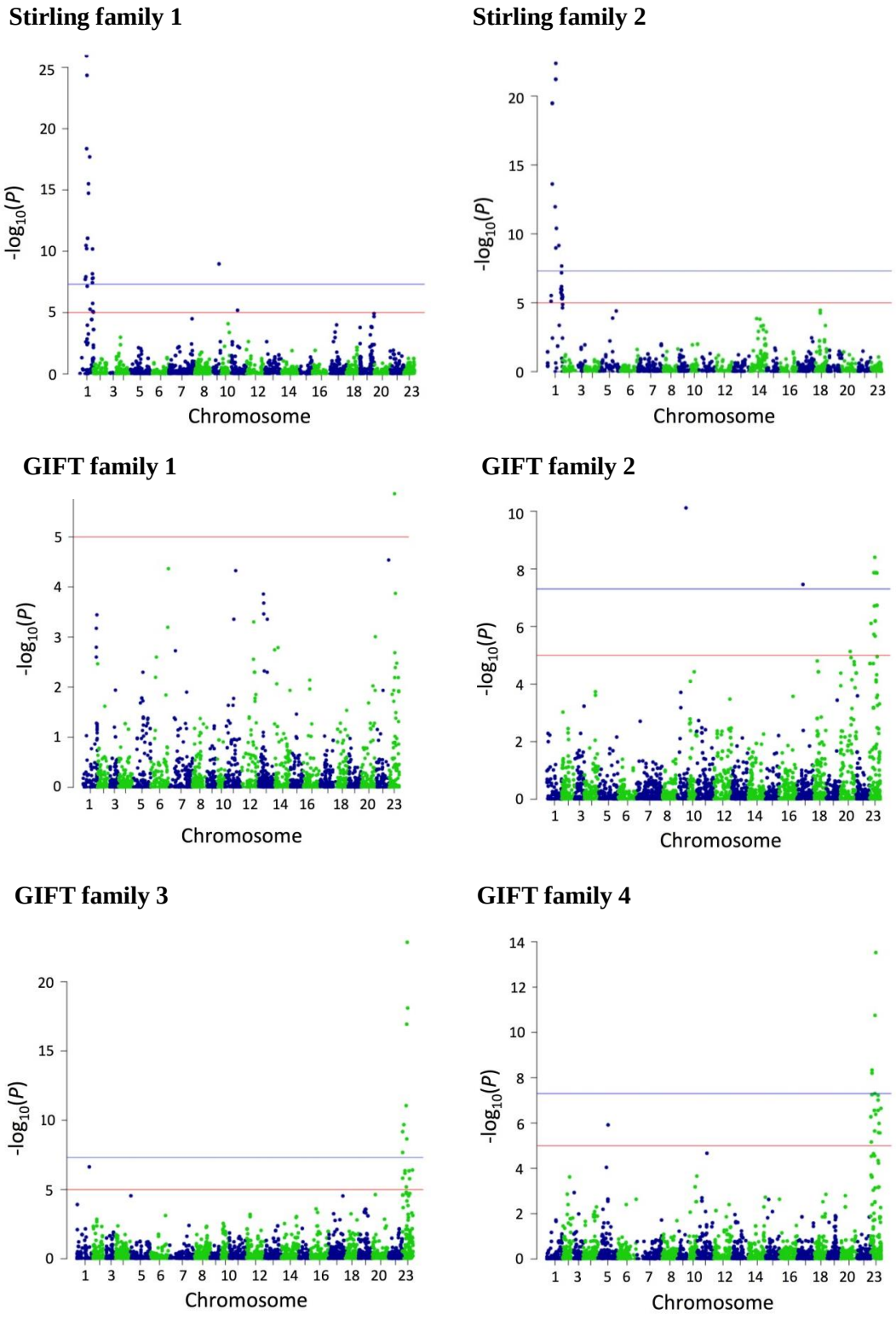
				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
19	2F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	5F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	6F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	13F105	Female	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	14F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	15F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	17F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	18F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	19F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	21F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	22F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	25F105	Female	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	27F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	28F105	Female	X/X	C/C		C/C	A/G	C/C	259/269	246/276	
19	31F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	33F105	Female	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	37F105	Female	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	38F105	Female	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	39F105	Female	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	3M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	4M105	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	7M105	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	8M105	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
19	9M105	Male	X/Y	C/T		C/T	A/G	C/C	267/275	244/274	
19	10M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	11M105	Male	X/Y	C/T		C/C	A/G	C/C	267/269	274/276	
19	12M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	16M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	20M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	23M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	24M105	Male	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	26M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	29M105	Male	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	30M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	32M105	Male	X/Y	C/T		C/C	A/G	C/T	267/269	244/274	
19	34M105	Male	X/X	C/C		C/C	G/G	C/T	259/275	244/246	
19	35M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
19	36M105	Male	X/X	C/C		C/C	G/G	C/T	259/269	246/276	
19	40M105	Male	X/Y	C/T		C/T	A/G	C/C	267/269	274/276	
	GIFT broodstock ID										
	5S	Male	X/Y	C/T	A/T	C/C	G/G	C/C	267/269	274/276	253/256
	6S	Male	X/Y	C/T	A/A	C/C	A/G	C/C	267/275	244/274	253/256
	8S	Male	X/Y	C/T	A/A	C/C	A/G	C/C	267/301	274/336	253/256
	10S	Male	X/Y	C/T	A/T	C/C	A/G	C/C	267/269	274/276	253/256
	20S	Male	X/X	C/C	A/A	C/C	A/G	C/C	251/275	244/284	253/256

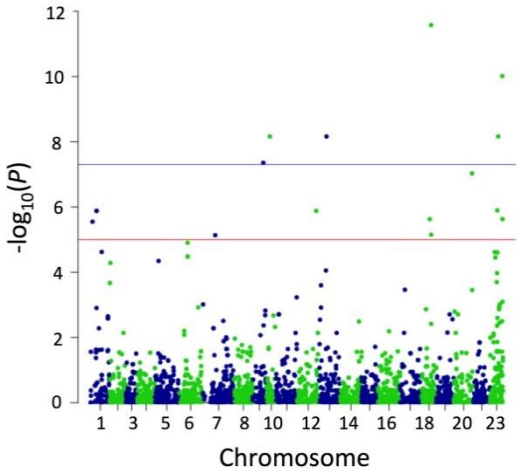
				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
	11S	Male	X/Y	C/T	A/T	C/C	A/A	C/C	267/269	274/276	253/256
	25S	Male	X/Y	C/T	A/T	C/C	G/G	C/C	267/269	274/276	253/256
	12S	Male	X/Y	C/T	A/T	C/C	A/G	C/C	267/269	274/276	253/256
	21S	Male	X/Y	C/T	A/T	C/C	A/G	C/C	267/269	274/276	253/256
	13S	Male	X/Y	C/C	A/T	C/C	G/G	C/C	267/269	274/276	253/256
	14S	Male	X/Y	C/C	A/T	C/C	G/G	C/C	267/269	274/276	253/256
	26S	Male	X/Y	C/T	A/A	C/C	A/A	C/C	267/275	244/274	253/256
	15S	Male	X/Y	C/T	A/A	C/C	A/G	C/C	251/267	274/284	256/256
	22S	Male	X/Y	T/T	A/A	C/C	A/G	C/C	267/267	274/274	256/256
	27S	Male	X/X	C/C	A/A	C/C	A/G	C/C	275/301	244/336	253/253
	16S	Male	X/Y	C/T	A/T	C/T	A/A	C/C	267/269	274/276	253/256
	28S	Male	X/Y	C/T	A/T	C/C	A/A	C/C	267/269	274/276	253/256
	17S	Male	X/Y	C/T	A/T	C/C	A/A	C/C	267/269	274/276	253/256
	23S	Male	X/Y	C/T	A/T	C/C	G/G	C/C	267/269	274/276	253/256
	18S	Male	X/Y	C/T	A/T	C/C	A/G	C/C	267/269	274/276	253/256
	5D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	6D	Female	X/X	C/C	A/T	C/C	G/G	C/C	269/301	276/336	253/253
	8D	Female	X/X	C/C	A/A	C/C	A/A	C/T	259/301	246/340	253/253
	9D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	10D	Female	X/X	C/C	T/T	C/C	G/G	C/C	269/279	276/276	253/253
	20D	Female	X/X	C/C	A/T	C/C	A/G	C/C	269/269	276/276	253/253
	11D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	12D	Female	X/X	C/C	A/T	C/C	G/G	C/T	259/269	246/276	253/253

				SNPs					Microsatellites		
Family	**ID**	**Sex**	***Amh* exon VII/ *Amh* exon VI**	***ss2017360173***	***ss2017360175***	***ss2017360134***	**ss2017360178**	***ss2017360168***	***UNH898***	***ARO172***	***Amh*_E0_del**
	21D	Female	X/X	C/T	A/A	C/C	A/G	C/C	251/285	252/284	253/256
	13D	Female	X/X	C/C	A/T	C/C	A/A	C/C	269/269	244/276	253/253
	14D	Female	X/X	C/T	A/T	C/C	G/G	C/C	269/285	252/276	253/253
	15D	Female	X/X	C/T	A/A	C/C	A/G	C/C	269/301	276/340	253/253
	22D	Female	X/X	C/C	A/T	C/C	A/G	C/C	269/301	276/336	253/253
	24D	Female	X/X	C/C	A/A	C/C	A/G	C/C	275/301	244/336	253/253
	16D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	17D	Female	X/X	C/C	A/T	C/C	G/G	C/C	269/301	276/336	253/253
	23D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	18D	Female	X/X	C/C	T/T	C/C	A/G	C/C	269/269	276/276	253/253
	Stirling broodstock										
	C1PF	Dam	X/X	T/T	A/A	C/C	G/G	C/T	259/291	246/274	253/253
	C1PM	Sire	X/X	T/T	A/A	C/C	G/G	C/T	259/279	246/272	253/253
	C2PF	Dam	X/X	T/T	A/A	C/C	G/G	C/T	259/271	246/246	253/253
	C2PM	Sire	X/X	T/T	A/A	C/C	G/G	C/T	259/279	246/272	253/253

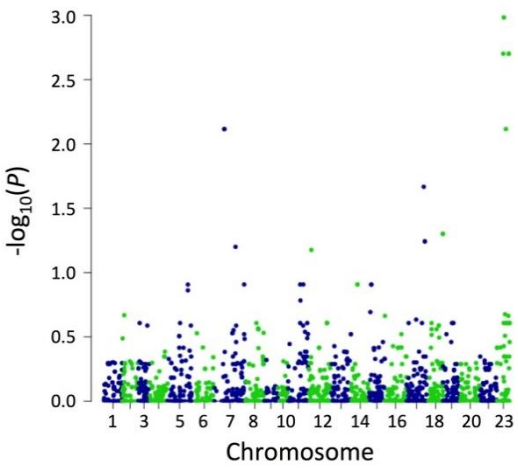
Figure C4.1 Genome-wide association plot with the phenotypic sex for each Stirling and GIFT family from BSA-ddRAD analysis. Each dot represents a SNP and the Y-axis represents the magnitude of association ($-\log_{10}P$ value of F-test) of the SNP with phenotypic sex, while the X-axis represents the position in the linkage groups of the assembled Nile tilapia genome. The alternating blue and green colours are used to distinguish between chromosomes. The red solid line represents a *q*-value of 0.05 and the blue solid line represents a *q*-value of 0.01.



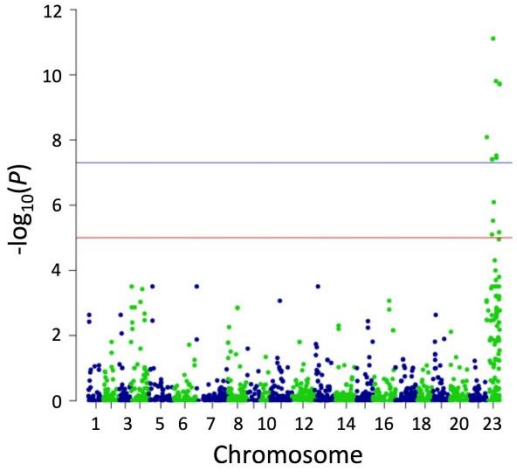
GIFT family 5



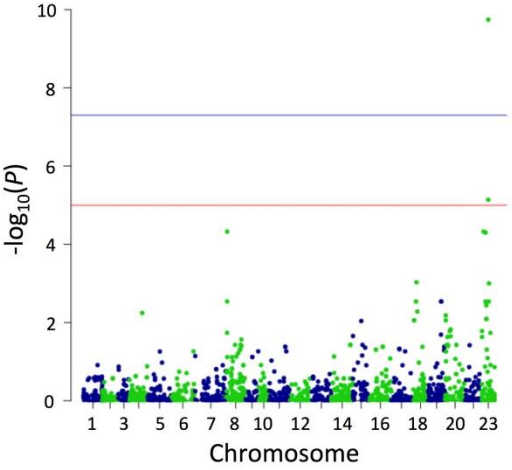
GIFT family 6



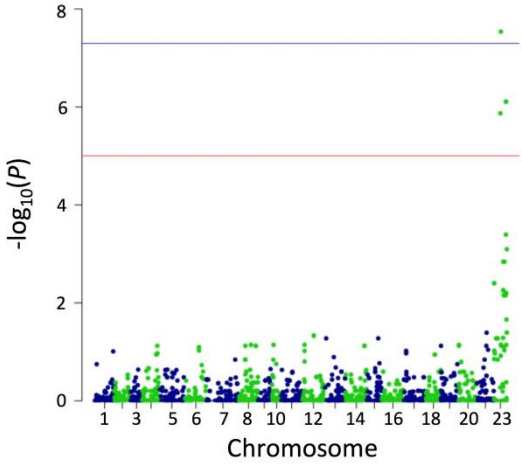
GIFT family 7



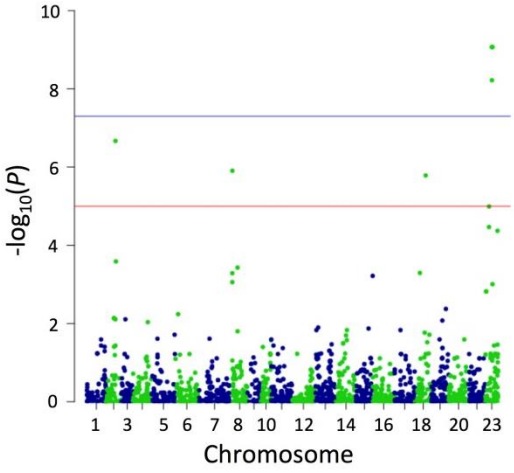
GIFT family 8



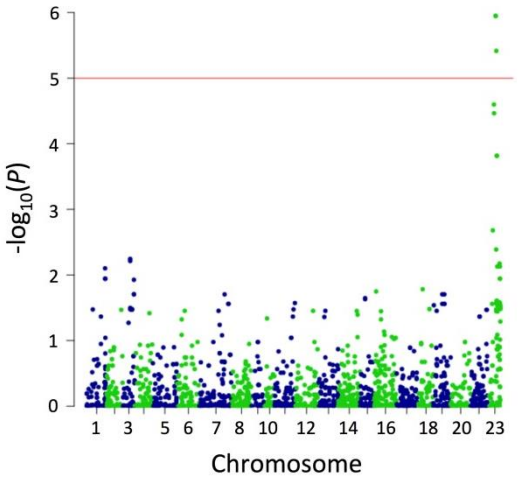
GIFT family 9



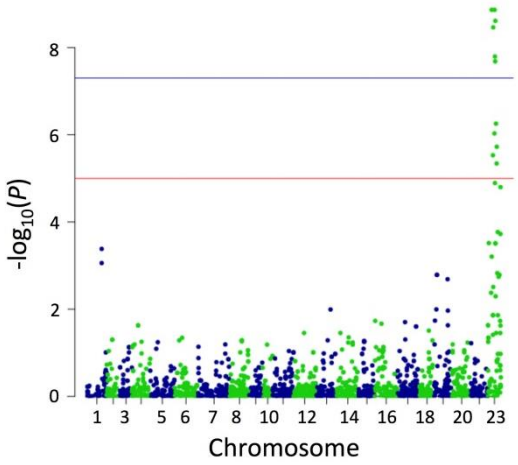
GIFT family 10



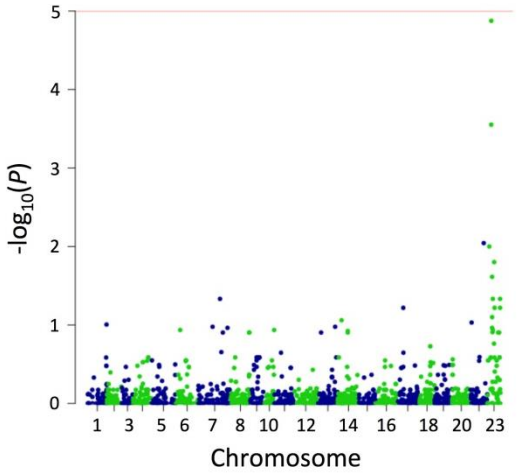
GIFT family 11



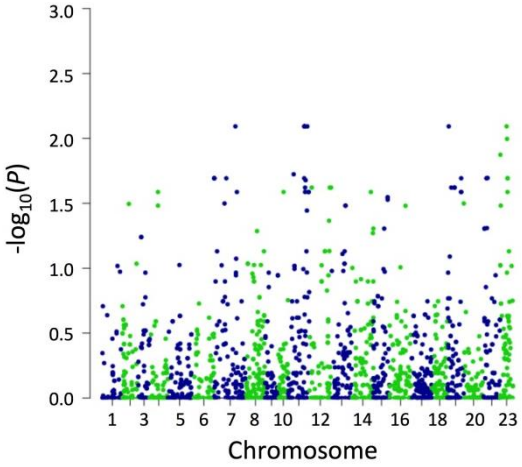
GIFT family 12



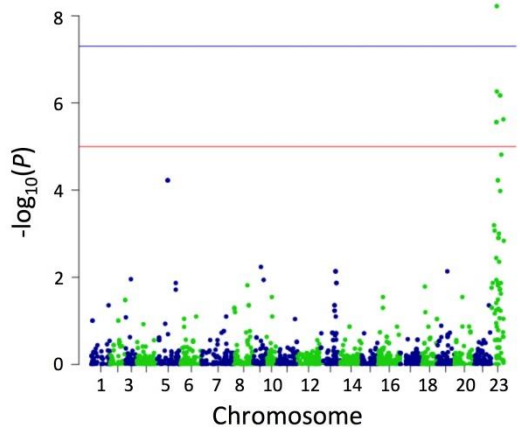
GIFT family 13



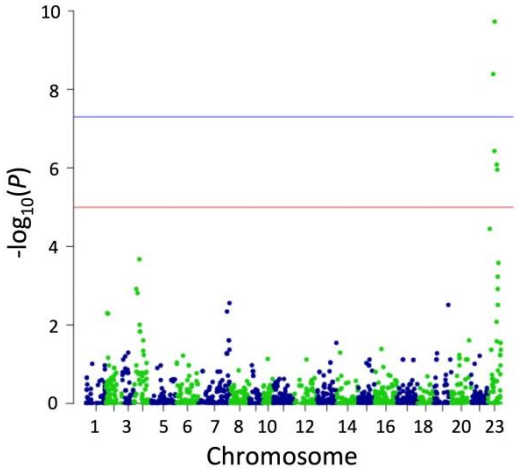
GIFT family 14



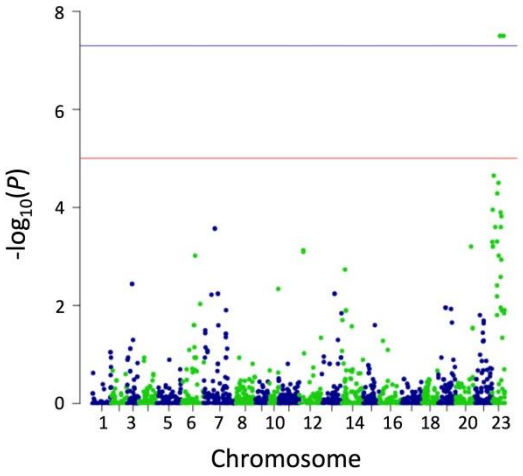
GIFT family 15



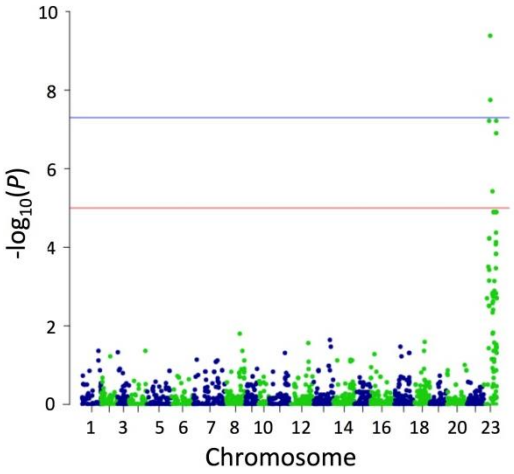
GIFT family 16



GIFT family 17



GIFT family 18



GIFT family 19

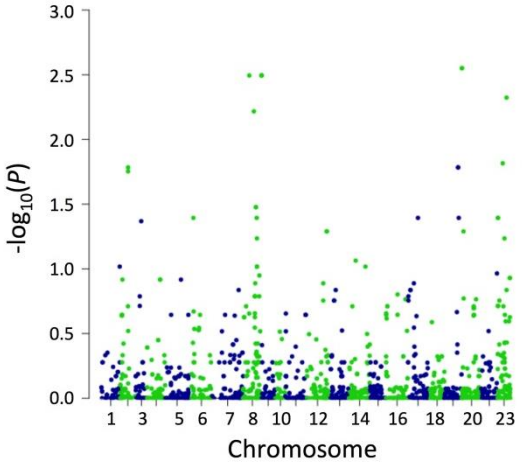
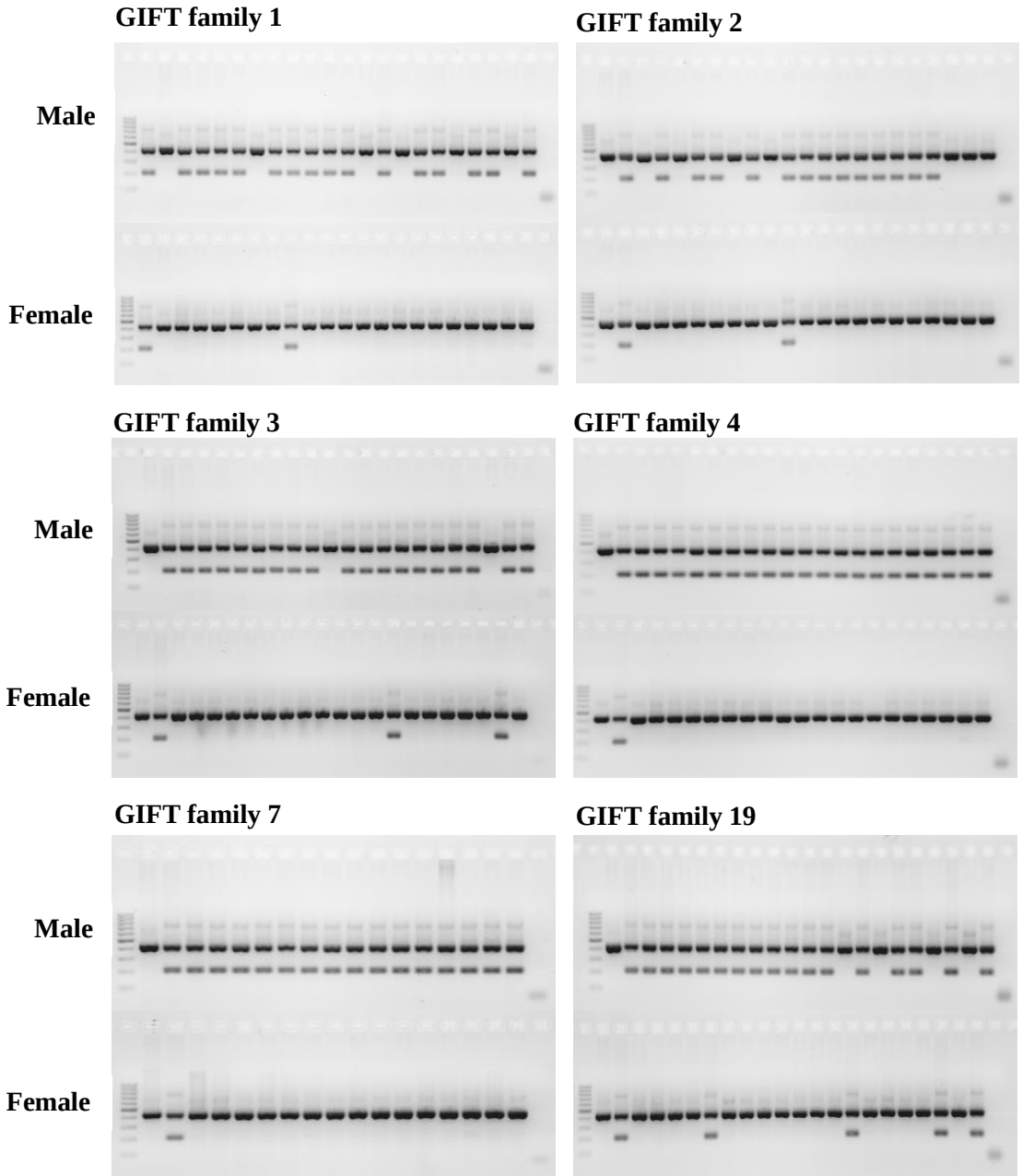


Figure C4.2 Amplified PCR products using *Amh* exon VII deletion (*Amhy*, Eshel *et al*. 2014) on 1.5 % agarose gel for six GIFT families individually. Nearly all males have a 233 bp deletion. The first lane is for male progenies whereas the lower lane is for female progenies for each gel picture. Each lane starts with 100 bp molecular marker, dam, sire and progenies and ends with a distilled water (no DNA) control.



Chapter 5

Table C5.1 Genotypes of the SNP markers in LG1 (*Oni23063*) and LG20 (*Oni3161*) for the Stirling broodstock.

Male (Wild/Red)					Female (Wild/Red)				
			LG1	LG20				LG1	LG20
Sample ID	**Sex**	**Type**	***Oni23063***	***Oni3161***	**Sample ID**	**Sex**	**Type**	***Oni23063***	***Oni3161***
1	Male	Wild	A/G	C/C	1	Female	Wild	G/G	T/T
2	Male	Wild	A/G	C/C	2	Female	Wild	G/G	C/T
3	Male	Wild	A/G	T/T	3	Female	Wild	G/G	C/C
4	Male	Wild	G/G	T/T	4	Female	Wild	G/G	C/T
5	Male	Wild	A/G	T/T	5	Female	Wild	G/G	T/T
6	Male	Wild	G/G	C/T	6	Female	Wild	G/G	T/T
7	Male	Wild	A/G	C/C	7	Female	Wild	G/G	C/T
8	Male	Wild	A/G	C/T	8	Female	Wild	G/G	C/T
9	Male	Wild	A/G	T/T	9	Female	Wild	G/G	C/T
10	Male	Wild	G/G	T/T	10	Female	Wild	G/G	C/T
11	Male	Wild	G/G	T/T	11	Female	Wild	G/G	C/C
12	Male	Wild	A/G	C/T	12	Female	Wild	G/G	T/T
13	Male	Wild	A/G	T/T	13	Female	Wild	G/G	C/T
14	Male	Wild	G/G	T/T	14	Female	Wild	G/G	T/T
15	Male	Wild	G/G	T/T	15	Female	Wild	G/G	T/T
16	Male	Wild	A/G	C/T	1	Female	Red	G/G	C/T
17	Male	Wild	A/G	C/C	2	Female	Red	G/G	T/T
18	Male	Wild	A/G	C/C	3	Female	Red	G/G	T/T
1	Male	Red	A/G	T/T	4	Female	Red	G/G	C/C
2	Male	Red	A/G	T/T	5	Female	Red	G/G	C/C
3	Male	Red	G/G	C/T	6	Female	Red	G/G	T/T
4	Male	Red	A/G	C/T	7	Female	Red	G/G	C/C
5	Male	Red	A/G	C/T	8	Female	Red	G/G	C/T
6	Male	Red	A/G	T/T	9	Female	Red	G/G	C/T
7	Male	Red	A/G	C/T	10	Female	Red	G/G	T/T
8	Male	Red	A/G	C/T	11	Female	Red	G/G	T/T
9	Male	Red	G/G	C/T	12	Female	Red	G/G	C/T
10	Male	Red	A/G	C/T	1	Clonal	Red	G/G	C/C
11	Male	Red	A/G	C/T	2	Clonal	Red	G/G	C/C
12	Male	Red	G/G	C/T					
13	Male	Red	A/G	C/T					
14	Male	Red	A/G	C/T					
15	Male	Red	A/G	C/T					
16	Male	Red	A/A	T/T					
17	Male	Red	A/A	T/T					

Chapter 6

Table C6.1 Information about each Nile tilapia broodstocks used in Chapter 6 with their tag number, phenotypic sex, strain of the broodstocks and genotypic sex for SNPs (*Oni23063*, *Oni28137*) and microsatellite (*UNH995*) markers in LG1 and SNP marker in LG20 (*Oni3161*).

			LG1			**LG20**
			SNPs		**Microsatellite**	**SNP**
Tag no.	**Phenotypic sex**	**Type**	***Oni23063***	***Oni28137***	***UNH995***	***Oni3161***
00068CFF63	Male	Wild	A/G	G/T	188/192	C/C
00068CF545	Male	Wild	A/G	G/T	184/188	C/C
00068CE677	Male	Wild	A/G	G/T	188/224	T/T
00068CD79E	Male	Wild	G/G	T/T	224/224	T/T
00068D074D	Male	Wild	A/G	G/T	188/224	T/T
00068D0624	Male	Wild	G/G	T/T	192/224	C/T
00068CF548	Male	Wild	A/G	G/T	188/192	C/C
00068CF4F8	Male	Wild	A/G	G/T	188/188	C/T
00068CEBBF	Male	Wild	A/G	G/T	188/190	T/T
00068D072C	Male	Wild	G/G	T/T	188/224	T/T
00068D0554	Male	Wild	A/G	G/T	188/224	T/T
00068CF4E1	Male	Wild	G/G	T/T	184/188	T/T
00068D009B	Male	Wild	A/G	G/T	184/188	C/T
00068D044F	Male	Wild	A/G	G/T	188/190	T/T
00068CFB49	Male	Wild	G/G	T/T	184/184	T/T
00068CEC9A	Male	Wild	G/G	T/T	192/224	T/T
00068CE248	Male	Wild	A/G	G/T	188/192	C/T
00068CEBC8	Male	Wild	A/G	G/T	188/188	C/C
00064E45020	Male	Wild	A/G	G/T	184/188	C/C
00068CDB4B	Female	Wild	G/G	T/T	184/224	C/T
00068D0462	Female	Wild	G/G	T/T	184/184	C/C
00068CEB8E	Female	Wild	G/G	T/T	190/192	C/T
00068CFA17	Female	Wild	G/G	T/T	224/224	T/T
00068CE32C	Female	Wild	G/G	T/T	190/192	T/T
00068D036B	Female	Wild	G/G	T/T	192/224	C/T
00068CFF78	Female	Wild	G/G	T/T	188/224	C/T
00068CFD17	Female	Wild	G/G	T/T	184/188	C/T
00068D0E1A	Female	Wild	G/G	T/T	190/192	C/T
00068CF9F4	Female	Wild	G/G	T/T	184/190	C/C
00068CF473	Female	Wild	G/G	T/T	188/190	T/T
00068CF550	Female	Wild	G/G	T/T	192/224	T/T
00068D0FCF	Female	Wild	G/G	T/T	184/188	C/T
00068CDF08	Female	Wild	G/G	T/T	N/A	T/T
00068CF853	Female	Wild	G/G	T/T	188/190	T/T

			LG1			**LG20**
			SNPs		**Microsatellite**	**SNP**
Tag no.	**Phenotypic sex**	**Type**	***Oni23063***	***Oni28137***	***UNH995***	***Oni3161***
00068CE8E1	Male	Red	A/G	G/T	236/252	T/T
00068D0301	Male	Red	A/G	G/T	232/252	T/T
00068CFB00	Male	Red	G/G	T/T	184/184	C/T
00068CECB9	Male	Red	A/G	G/T	184/236	C/T
00068CF317	Male	Red	A/G	G/T	188/236	C/T
00068D0D4A	Male	Red	A/G	G/T	188/236	T/T
00068CE1CB	Male	Red	A/G	G/T	188/236	C/T
00068CE85C	Male	Red	A/G	G/T	184/236	C/T
00068CE1F9	Male	Red	G/G	T/T	184/188	C/T
00068CE652	Male	Red	A/G	G/T	188/188	C/T
00068CD668	Male	Red	A/G	G/T	184/236	C/T
00068CDF8A	Male	Red	G/G	T/T	184/184	C/T
00068CFADA	Male	Red	A/G	G/T	184/236	C/T
00068CDC84	Male	Red	A/G	G/T	188/236	C/T
00068D0A91	Male	Red	A/G	G/T	184/236	T/T
00068CFB26	Male	Red	A/G	G/T	232/252	T/T
00068CE0C3	Male	Red	A/A	G/G	236/236	T/T
00068CEF4E	Male	Red	A/G	G/T	188/236	T/T
00068CF40D	Male	Red	A/G	G/T	188/236	C/T
00068CDC83	Male	Red	A/G	G/T	184/184	C/T
00068CED41	Male	Red	A/A	G/G	232/232	T/T
00068CE93A	Male	Red	A/A	G/G	232/232	T/T
0001E0A56A	Male	Red	A/G	G/T	184/236	C/T
00068CEA72	Female	Red	G/G	T/T	184/184	C/T
00068D0097	Female	Red	G/G	T/T	184/184	T/T
00068CEF72	Female	Red	G/G	T/T	184/188	T/T
00068CFFE3	Female	Red	G/G	T/T	184/184	C/C
00068CDE0C	Female	Red	G/G	T/T	184/184	C/C
00068D0EE9	Female	Red	G/G	T/T	184/188	C/T
00068CEE98	Female	Red	G/G	T/T	184/184	T/T
00068CE1D6	Female	Red	G/G	T/T	184/184	C/C
00068CE02D	Female	Red	G/G	T/T	184/184	C/T
00068D01BC	Female	Red	G/G	T/T	184/188	C/T
00068D005E	Female	Red	G/G	T/T	184/188	T/T
00068CF895	Female	Red	G/G	T/T	184/188	T/T

Chapter 7

Table C7.1 Details of the 22 putative species-diagnostic SNP markers with their genotypes for *O. mossambicus, O. niloticus* and *O. aureus*, linkage group (LG), chromosomal position, DNA strand and SNP position (Syaifudin 2015).

Marker ID	Species	*O. niloticus*	*O. aureus*	*O. mossambicus*	LG	Position	Strand	SNP position
Omos2007	*O. mossambicus*	GG	GG	AA	LG13	9212591	-	52
Omos2657	*O. mossambicus*	GG	GG	AA	LG15	11282183	+	95
Omos3481	*O. mossambicus*	TT	TT	CC	LG16_21	33845307	+	41
Omos7956	*O. mossambicus*	CC	CC	TT	LG4	14941897	-	34
Omos10120	*O. mossambicus*	GG	GG	AA	LG7	40145502	-	36
Omos10818	*O. mossambicus*	GG	GG	AA	LG8_24	3390946	+	128
Omos3582	*O. mossambicus*	GG	GG	AA	LG16_21	563234	-	55
Omos8084	*O. mossambicus*	AA	AA	GG	LG4	22474196	-	8
Omos4092	*O. mossambicus*	TT	TT	CC	LG17	3619802	+	18
Onil2675	*O. niloticus*	GG	TT	TT	LG15	12494644	-	6
Onil3057	*O. niloticus*	AA	TT	TT	LG15	909826	-	14
Onil5782	*O. niloticus*	AA	TT	TT	LG20	16532929	-	66
Onil9497	*O. niloticus*	TT	CC	CC	LG6	7648628	-	79
Onil1276	*O. niloticus*	AA	GG	GG	LG12	3072836	-	128
Oaur966	*O. aureus*	CC	TT	CC	LG12	13862286	+	100
Oaur9418	*O. aureus*	TT	CC	TT	LG6	36351716	+	27
Oaur8029	*O. aureus*	AA	GG	AA	LG4	2024866	-	21
Oaur3001	*O. aureus*	GG	AA	GG	LG15	6075396	+	21
Oaur2890	*O. aureus*	CC	TT	CC	LG15	24160607	-	129
Oaur3873	*O. aureus*	TT	CC	TT	LG17	21279151	-	29
Oaur5416	*O. aureus*	CC	TT	CC	LG1	27201771	+	36
Oaur4411	*O. aureus*	CC	TT	CC	LG18	20048688	-	85

Table C7.2 KASP assay sequences used in Chapter 7.

Assay ID	Primer_Allele FAM	Primer_Allele HEX	Primer_Common	Allele FAM	Allele HEX
Omos2007	GAAGGTGGCCGATATG GTAGCTT	AAGGTGGCCGATATGG TAGCTC	CTGCATTAGAAATKAG AGCTGTTTGGCTT	A	G
Omos2657	ACCATCAATGCTGAAA GATACTGACAA	CCATCAATGCTGAAAG ATACTGACAG	GAATGGGAGCATATG CTGCTCTGAA	A	G
Omos3481	AATTTGGCATAAATGA AGCTTCCTTAAAC	GAATTTGGCATAAATGA AGCTTCCTTAAAT	CAAAGCTATGAAACC ATTAATGGGTCAATT	C	T
Omos7956	ATACAACTACATTTAGC CAAACTTCTAAC	GATACAACTACATTTAG CCAAACTTCTAAT	CCAAGGGATCATGTGG GGATATCAA	C	T
Onil2675	CACTATGCTCATCCTGC AGGG	GTCACTATGCTCATCCT GCAGGT	GGCCAGTGAGCACACT TAATTGGAA	G	T
Onil3057	GTTAACAGTGGTGTCCT ACATCAAATAT	GTTAACAGTGGTGTCCT ACATCAAATAA	AGACGTCATAATGCCA CGCATGCAT	A	T
Onil5782	CCTAAACCATAAAGTT GGGAGGATGT	CCTAAACCATAAAGTTG GGAGGATGA	TGTCAGATGCAGCACA CMAAGACATTT	A	T
Onil9497	GTTCAAAGTGCTTATCA TGAGCTCG	GGTTCAAAGTGCTTATC ATGAGCTCA	CTCCGGAAAACGGCTA CATAGGTA	C	T
Oaur966	CATATGCAAATTAATTG ACACGGACC	GCATATGCAAATTAATT GACACGGACT	GAAAAGACCAGCATA TGGGGAGGAA	C	T
Oaur9418	CAATCTAAGGCCAAGA GTCTCAG	CCAATCTAAGGCCAAG AGTCTCAA	CATGCTATACCAACCT TGAGGCTGTT	C	T
Oaur5416	TTCACAATCACCCAGC ACCG	GCTTTCACAATCACCCA GCACCA	CATGCCGGAGATCAGC AATGAAGAT	C	T
Omos10818	GTTCAGCAAGCTTTCAA CCTGCC	GTTCAGCAAGCTTTCAA CCTGCT	TCACCATGCAAGACTC CATTAATGAGAAA	G	A

Table C7.3 Details of each sample used for ddRADseq: sample ID, sequencing run, P1 and P2 barcode information and the generated paired-end reads.

Sample ID	Run	P1 Barcode	P2 Barcode	Paired-end reads
09Dam_1	1	TCAGA	TAGCA	123643
09Sire_1	1	GATCG	TAGCA	157298
12Dam_1	1	TCGAG	TAGCA	126529
12Sire_1	1	GTCAC	TAGCA	121730
13Dam_1	1	TGCAACA	TAGCA	142380
13Sire_1	1	CGTATCA	TAGCA	190345
15Dam_1	1	TCTCTCA	AGCTGTC	111468
15Sire_1	1	GTACACA	AGCTGTC	137183
23Dam_1	1	ACGTA	AGCTGTC	189657
23Sire_1	1	AGAGT	AGCTGTC	160752
09Dam_2	1	TCTCTCA	TACGTGT	81705
9Sire_2	1	GTACACA	TACGTGT	86160
12Dam_2	1	ACGTA	TACGTGT	113777
12Sire_2	1	AGAGT	TACGTGT	203213
13Dam_2	1	CAGTCAC	TACGTGT	99497
13Sire_2	1	GCTAACA	TACGTGT	131533
15Dam_2	1	TCAGA	GCATA	84256
15Sire_2	1	GATCG	GCATA	81829
23Dam_2	1	TCGAG	GCATA	100642
23Sire_2	1	GTCAC	GCATA	87363
09Dam_3	1	TCAGA	CATCTGT	157346
9Sire_3	1	GATCG	CATCTGT	173931
12Dam_3	1	TCGAG	CATCTGT	202028
12Sire_3	1	GTCAC	CATCTGT	176908
13Dam_3	1	TGCAACA	CATCTGT	175317
13Sire_3	1	CGTATCA	CATCTGT	233969
15Dam_3	1	TCTCTCA	CTGGT	107502
15Sire_3	1	GTACACA	CTGGT	109129
23Dam_3	1	ACGTA	CTGGT	137201
23Sire_3	1	AGAGT	CTGGT	119525
025Dam	2	TCAGA	GCATA	112012
025Sire	2	TGCAACA	GAGATGT	153459
027Dam	2	CATGA	CTGGT	143505
027Sire	2	CACAGAC	GTCAAGT	188324
031Dam	2	TCGAG	TAGCA	180441
031Sire	2	TCTCTCA	AGCTGTC	147983
033Dam	2	GCATT	GCATA	161261
037Dam	2	CTAGGAC	CATCTGT	248577
037Sire	2	ACGTA	CTGGT	278181
044Dam	2	GCTAACA	ATACGGT	144088
044Sire	2	ATGCT	TAGCA	114381
050Dam	2	AGGACAC	TACGTGT	192117
050Sire	2	TCAGA	GAGATGT	228102
052Dam	2	CGTATCA	CTGGT	188587
052Sire	2	CATGA	GTCAAGT	245158

Sample ID	Run	P1 Barcode	P2 Barcode	Paired-end reads
053Dam	2	ACTGCAC	TAGCA	179638
053Sire	2	TCGAG	AGCTGTC	194859
056Dam	2	GTACACA	GCATA	176234
056Sire	2	GCATT	GAGATGT	203256
087Dam	2	CTAGGAC	CTGGT	228579
087Sire	2	ACGTA	GTCAAGT	207664
097Dam1	2	GCTAACA	TAGCA	301583
097Dam2	2	ATGCT	AGTCA	192209
097Sire1	2	ATGCT	AGCTGTC	183624
097Sire2	2	ACACGAG	TACGTGT	177809
101Dam	2	AGGACAC	GCATA	141976
101Sire	2	TCAGA	CGATC	178195
105Dam	2	CGTATCA	GTCAAGT	173236
105Sire	2	CATGA	GAAGC	259019
040Dam	2	ACTGCAC	AGCTGTC	198548
051Dam	2	TCGAG	AGTCA	289456
058Dam	2	TCTCTCA	TACGTGT	187082
059Dam	2	GTCAC	GCATA	171409
099Dam1	2	GTACACA	GAGATGT	225223
099Dam2	2	GACTA	GCATA	172852
040Sir	2	GCATT	CGATC	152076
048Sir	2	CTCTTCA	CATCTGT	177146
051Sir	2	CGATA	CTGGT	185560
055Sir	2	CTAGGAC	GTCAAGT	117645
058Sir	2	ACGTA	GAAGC	230767
080Sir	2	CAGTCAC	ATACGGT	226247
092Sir	2	AGAGT	TAGCA	189689
099Sir1	2	GCTAACA	AGCTGTC	123326
099Sir2	2	AGGACAC	GAGATGT	123157