The genetic and epigenetic architecture of sexual dimorphism in tilapia

Wan, Zi Yi

2019

Wan, Z. Y. (2019). The genetic and epigenetic architecture of sexual dimorphism in tilapia.
Doctoral thesis, Nanyang Technological University, Singapore.

<https://hdl.handle.net/10356/136866>

<https://doi.org/10.32657/10356/136866>

This work is licensed under a Creative Commons Attribution-NonCommercial 4.0
International License (CC BY-NC 4.0).

Downloaded on 26 Apr 2023 22:05:51 SGT

**NANYANG
TECHNOLOGICAL
UNIVERSITY**

SINGAPORE

The Genetic and Epigenetic Architecture of
Sexual Dimorphism in Tilapia

WAN ZI YI

SCHOOL OF BIOLOGICAL SCIENCES

2019

The Genetic and Epigenetic Architecture of
Sexual Dimorphism in Tilapia

WAN ZI YI

SCHOOL OF BIOLOGICAL SCIENCES

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

2019

Statement of Originality

I hereby certify that the work embodied in this thesis is the result of original research done by me except where otherwise stated in this thesis. The thesis work has not been submitted for a degree or professional qualification to any other university or institution. I declare that this thesis is written by myself and is free of plagiarism and of sufficient grammatical clarity to be examined. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

26/7/2019

.....
Date

.....
WAN ZI YI

Supervisor Declaration Statement

I have reviewed the content and presentation style of this thesis and declare it of sufficient grammatical clarity to be examined. To the best of my knowledge, the thesis is free of plagiarism and the research and writing are those of the candidate's except as acknowledged in the Author Attribution Statement. I confirm that the investigations were conducted in accord with the ethics policies and integrity standards of Nanyang Technological University and that the research data are presented honestly and without prejudice.

26/7/2019

Date

Assoc Prof Yue Genhua

28/7/2019

Date

Assoc Prof Lin Chun Ling Valerie

Authorship Attribution Statement

This thesis contains material from 2 papers published in the following peer-reviewed journal(s) and 1 poster presentation accepted at the conference in which I am listed as an author.

The results on whole genome bisulfite sequencing of tilapia muscle tissue were published as **Zi Yi Wan, Jun Hong Xia, Grace Lin, Le Wang, Valerie CL Lin, and Gen Hua Yue. "Genome-wide methylation analysis identified sexually dimorphic methylated regions in hybrid tilapia." *Scientific Reports* 6 (2016): 35903.**

The contributions of the co-authors are as follows:

- A/Prof GH Yue and A/Prof VCL Lin conceived the study.
- I prepared the manuscript drafts. The manuscript was revised by Jun Hong Xia, Grace Lin, Le Wang, Valerie CL Lin, and Gen Hua Yue.
- I co-designed the study with Jun Hong Xia and Grace Lin and performed all the laboratory work at the Temasek Life Sciences Laboratory.
- I performed all the bioinformatics analysis using the high-performance computing clusters at the Temasek Life Sciences Laboratory.
- Next-generation sequencing service was provided by BGI-Shenzhen, China.

The results on muscle transcriptome of sex-biased genes in tilapia were published as **Zi Yi Wan, Grace Lin, and Gen Hua Yue. "Genes for sexual body size dimorphism in hybrid tilapia (*Oreochromis sp. x Oreochromis mossambicus*)."** *Aquaculture and Fisheries* (2019). (Online)

The contributions of the co-authors are as follows:

- A/Prof GH Yue conceived the study.
- I prepared the manuscript drafts. The manuscript was revised by Grace Lin and A/Prof Gen Hua Yue.
- I co-designed the study with Grace Lin and performed all the laboratory work at the Temasek Life Sciences Laboratory.
- I performed all the bioinformatics analysis using the high-performance computing clusters at the Temasek Life Sciences Laboratory.
- Next-generation sequencing service was provided by BGI-Shenzhen, China.

The results on brain transcriptome of sex-biased genes in tilapia were presented as a poster presentation in the International Symposium of Genetics in Aquaculture ISGA XIII (15 - 20 July 2018), Cairns, Australia, as **Zi Yi Wan, Le Wang, Valerie CL Lin, and Gen Hua Yue. Transcriptomes of the sexually dimorphic central nervous system of developing Mozambique tilapia (*Oreochromis mossambicus*).**

The contributions of the co-authors are as follows:

- A/Prof Gen Hua Yue and A/Prof Valerie Chun Ling Lin conceived the study.
- I co-designed the study with A/Prof Gen Hua Yue and performed all the laboratory work at the Temasek Life Sciences Laboratory.
- I performed all the bioinformatics analysis using the high-performance computing clusters at Temasek Life Sciences Laboratory.
- Next-generation sequencing was conducted in Temasek Life Sciences Laboratory using an Illumina NextSeq 500.

26/7/2019

Date

WAN ZI YI

Acknowledgements

First and foremost, I would like to express my sincere gratitude to Dr Yue Gen Hua, for encouraging me to take up a PhD course and for his supervision and support throughout the entire project. Without his support, I would not have overcome the difficulties faced in finishing this project. I would also like to thank Prof Valerie, for being my co-supervisor and provided valueable suggestions in some parts of the project. I am also very grateful to the members of my Thesis Advisory Committee Prof Melissa Fullwood and Prof Christoph Winkler for their valueable advice. I would also like to thank my lab members, past and present, for their help throughout the course of my PhD. Special thanks to Dr Liu Peng and Dr Wang Le, for initiating me into bioinformatics, Bao Qing for statistical analysis support, May and Sun Fei for their help in some of my experiments. These works would also not be possible without the help and guidance from the bioinformatic experts on Biostars.org.

Last but not least, I would like to thank my mom and dad for their unwavering support throughout the course of my PhD. Special thanks to my wife, Sze Chee, for her constant encouragement and support. You have been my constant source of motivations.

Abbreviations

4E-BP1	Initiation Factor 4E-Binding Protein 1
5mC	Fifth Carbon Of A Cytosine Base
ACR	Active Cis-Regulatory Region
AgRP	Agouti-Related Protein
Akt	Protein Kinase B
AMHy	Anti-Müllerian Hormone Y
BAM	Binary Alignment/Map Format
BLAST	Basic Local Alignment Search Tool
BMP7	Bone Morphogenic Protein 7
Bp	Base Pair
cDNA	Complementary DNA
CHG	5'-C-A/T/G-G-3' Cytosine Methylation Site
CHH	5'-C-A/T/G-A/T/G-3' Cytosine Methylation Site
CHSE	Chse-214 Cell Line From Salmon
CpG	5'-C-Phosphate-G-3' Cytosine Methylation Site
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
ddRADseq	Double Digest Restriction-Site Associated DNA Sequencing
DEG	Differentially Expressed Gene
DMR	Differentially Methylated Region
<i>dN/dS</i>	Non-Synonymous To Synonymous Substitution Rate
DNA	Deoxyribonucleic Acid
DPH	Days Post Hatch
EGFR	Epidermal Growth Factor Receptor
ENU	N-Ethyl-N-Nitrosourea
ERK	Extracellular Signal-Regulated Kinase
ESRRA	Estrogen-Related Receptor Alpha
FBS	Fetal Bovine Serum
FCR	Feed Conversion Ratio
FDR	False Discovery Rate
FISH	Fluorescence In Situ Hybridization
FOXO4	Forkhead Box O4
GFF	General Feature Format For Gene Annotation
GFP	Green Fluorescence Protein
GH	Growth Hormone
GHRH	Growth Hormone Releasing Hormone
GIFT	Genetically Improved Farmed Tilapia
GnRH	Gonadotrophin Releasing Hormone
GREB1	Growth Regulation By Estrogen In Breast Cancer 1
gRNA	Guide RNA
GWAS	Genome-Wide Association Studies
IGF1	Insulin-Like Growth Factor 1

IGF1R	Insulin-Like Growth Factor Receptor 1
INDEL	Insertion-Deletion
IRS1	Insulin Receptor Substrate 1
IRS2	Insulin Receptor Substrate 2
LB	Luria Broth
LEF1	Lymphoid Enhancer-Binding Factor-1
LG	Linkage Group
LINE	Long Interspersed Nuclear Element
LMR	Low-Methylated Region
LPIN1	Lipin 1
LTR	Long Terminal Repeat
MAPK	Mitogen-Activated Protein Kinase
MEK	Mitogen-Activated Protein Kinase Kinase
min	Minute
M-MLV	Moloney Murine Leukemia Virus
mRNA	Messenger Ribonucleic Acid
mTOR	Mammalian Target Of Rapamycin
mTORC1	Mammalian Target Of Rapamycin Complex 1
NGS	Next Generation Sequencing
NHEJ	Non-Homologous End Joining
NPY	Neuropeptide Y
OmB	Oreochromis Mossambicus Brain Cell Line
PAM	Protospacer Adjacent Motif
PAML	Phylogenetic Analysis By Maximum Likelihood
PBS	Phosphate-Buffered Saline
PCR	Polymerase Chain Reaction
PI3K	Phosphoinositide 3-Kinase
PMD	Partially Methylated Domain
POMC	Pro-Opiomelanocortin
qPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Traits Loci
RASGRF1	Ras Protein-Specific Guanine Nucleotide-Releasing Factor 1
RNA	Ribonucleic Acid
RNA-seq	RNA Sequencing
RNP	Ribonucleoprotein
RPKM	Reads Per Kilo Million
RPM	Revolution Per Minute
rRNA	Ribosomal RNA
SAM	Sequence Alignment/Map Format
sgRNA	Single Guide RNA
SINE	Short Interspersed Nuclear Element
SNP	Single Nucleotide Polymorphism
SPATA17	Spermatogenesis Associated 17

SRIF	Somastatin
SSR	Simple Sequence Repeat
TALEN	Transcription Activator-Like Effector Nuclease
TAp63	Transcriptionally Active p63
TE	Transposable Element
TM	Primer Melting Temperature
TSS	Transcriptional Start Site
TTS	Transcriptional Termination Site
UMR	Unmethylated Region
UTR	Untranslated Region
WGBS	Whole Genome Bisulfite Sequencing
ZFAND3	Zinc Finger AN1-Type Containing 3
ZFN	Zinc Finger Nuclease

List of Figures

Number	Figure Title	Page
Figure 1	Sexual dimorphism between female and male tilapia	6
Figure 2	Linkage map of chromosome LG1 in Mozambique tilapia	10
Figure 3	Project roadmap outlined for this thesis	37
Figure 4	Cartoon showing the phenotypic differences between male and female tilapia genital organs	40
Figure 5	Correlation matrix showing the Pearson correlation of base resolution CpG methylation genome-wide between the skeletal muscle tissue samples of hybrid tilapia	81
Figure 6	CpG Methylation level in relation to gene body	83
Figure 7	Distribution of CpG methylation level across various genomic elements such as promoters, introns, exons, DNA transposons, SINEs, LINEs and LTRs in male and female tilapia	87
Figure 8	CIRCOS chart summarizing the differentially methylated regions in two sexes of hybrid tilapia	90
Figure 9	Numbers of hypermethylated DMRs in male and female tilapia skeletal muscle tissue	91
Figure 10	Scatterplot showing relationship between numbers of methylated CpG and gene expression level	93
Figure 11	Representative illustration of identified active cis-regulatory region in the tilapia genome using whole genome bisulfite sequencing	95
Figure 12	Enrichments of sexually dimorphic active cis-regulatory motifs in hybrid tilapia skeletal muscles	97
Figure 13	Numbers of sex-biased genes on different chromosomes are shown on the bar plot	101
Figure 14	ClusterProfiler diagram summarizing the overrepresentations of gene ontologies in sex-biased genes identified in tilapia skeletal muscle	103
Figure 15	Non-synonymous to synonymous mutation rate ratio (dN/dS) of sex-biased genes in skeletal muscle	105
Figure 16	Haplotype of sex determining region for Mozambique tilapia at chromosome LG1	109
Figure 17	MA plot of 5 dph sex-biased genes in the tilapia brain	113
Figure 18	MA plot of 30 dph sex-biased genes in the tilapia brain	114
Figure 19	MA plot of 90 dph sex-biased genes in the tilapia brain	115
Figure 20	Chromosomal locations of observed and expected number of sex-biased genes in 5 dph tilapia	116
Figure 21	Chromosomal locations of observed and expected number of sex-biased genes in 30 dph tilapia	117
Figure 22	Chromosomal locations of observed and expected number of sex-biased genes in 90 dph tilapia	118
Figure 23	ClusterProfiler diagram summarizing the overrepresentations of gene ontologies in sex-biased genes identified in 5 dph tilapia brain tissue	119
Figure 24	ClusterProfiler diagram summarizing the overrepresentations of gene ontologies in sex-biased genes identified in 30 dph tilapia brain tissue	120
Figure 25	ClusterProfiler diagram summarizing the overrepresentations of gene ontologies in sex-biased genes identified in 90 dph tilapia brain tissue	121
Figure 26	The dN/dS ratio of sex-biased genes identified from tilapia brain transcriptomes	130
Figure 27	dN/dS ratio of sex-biased genes identified from tilapia brain in 0.1 intervals	131

Figure 28	Association of SNPs with the phenotype of sex	134
Figure 29	Expression level of <i>POMC</i> in the brain from 5 dph to 120 dph	138
Figure 30	Induction of luciferase activity in pomcERE-pGL3-promoter transfected OmB cells in ethanol treatment and β -Estradiol treatment at 48 hours post transfection	140
Figure 31	T7 genotyping assay of F0 fishes	142
Figure 32	Direct sequencing of PCR product amplified from 600 bp region flanking the target site	143
Figure 33	Estimated allele frequency of F0 <i>pomc</i> mutants based on T-vector clones	144
Figure 34	PCR genotyping of F1 mutants showing large deletions of approximately 380 bp and 1600 bp deletions	145
Figure 35	Expression level of <i>POMC</i> in wildtype zebrafish and <i>pomc</i> mutants	146
Figure 36	Anterior lateral view of zebrafish F0 mutants showing phenotype of small and less dense melanophores on F0 <i>pomc</i> mutants	147
Figure 37	Posterior lateral view of zebrafish F0 mutants showing phenotype of small and less dense melanophores on F0 <i>pomc</i> mutants	148
Figure 38	Quantitative analysis of melanophore distribution for the anterior and posterior trunk lateral view of F0 zebrafish mutants and wildtype control	149
Figure 39	Summary of phenotypic comparison between wildtype zebrafish and <i>pomc</i> zebrafish for traits of body weight and body length	151
Figure 40	The median difference in feeding latency between wildtype and <i>pomc-t3</i> zebrafish	153
Figure 41	Proposed mechanism of <i>POMC</i> in regulating sexual size dimorphism in tilapia	180

List of Tables

Table No	Table Title	Page
Table 1	Primers for genotyping microsatellite sex markers	42
Table 2	List of primers used for qPCR validations in skeletal muscle tissue	57
Table 3	List of primers for qPCR validations in brain tissue	58
Table 4	Primers for cloning POMC estrogen response elements	68
Table 5	Oligos used in generating template for T7 <i>in vitro</i> transcription of sgRNA for microinjection	70
Table 6	Primers flanking target sites of <i>POMC</i> sgRNA for genotyping assay	74
Table 7	Statistical summary of skeletal muscle whole genome bisulfite sequencing for all four libraries	80
Table 8	Summary of repeat elements masked with RepeatMasker based on vertebrate repeat element homology search in the tilapia genome	85
Table 9	Summary of RNA-seq statistics for skeletal muscle transcriptome	98
Table 10	Summary of RNA-seq statistics for the brain transcriptome 5 dph, 30 dph and 90 dph	110
Table 11	Summary of Panther Pathways enrichment test for sex-biased genes in 5 dph tilapia brain transcriptome	123
Table 12	Summary of Panther Pathways enrichment test for sex-biased genes in 30 dph tilapia brain transcriptome	124
Table 13	Summary of Panther Pathways enrichment test for sex-biased genes in 90 dph tilapia brain transcriptome	125

List of Supplementary Materials

Table No	Table title	Page
Supp. Table 1	Bodyweight to SNPs association summary in hybrid tilapia.	201
Supp. Table 2	Microsatellite genotype of sex-markers linked to sex locus.	202
Supp. Figure 1	Correlation between RNA-seq and RT-qPCR from muscle transcriptome.	204
Supp. Figure 2	Correlation between RNA-seq and RT-qPCR from brain transcriptome.	205

Contents

1	<i>Abstract</i>	1
2	<i>Introduction</i>	2
2.1	Sexual dimorphism and its prevalence in teleosts	2
2.2	Biology of tilapia	5
2.2.1	Overview of tilapia biology and tilapia as an aquaculture species	5
2.2.2	Sexual dimorphism in tilapia and its implication to aquaculture	6
2.2.3	Sex-determining loci in tilapia	8
2.2.4	Availability of genetic and genomic toolkits in tilapia	10
2.3	Current understanding of sexual size dimorphism in teleosts	11
2.3.1	Teleost growth regulation mechanisms	12
2.3.2	Teleost gonads have direct regulation on somatic growth __ Error! Bookmark not defined.	
2.3.3	Sex-biased gene expression in teleost somatic tissue	15
2.3.4	Differential CpG DNA methylation between sexes	17
2.3.5	Molecular evolution of sexual dimorphism	18
2.4	Overview of genomics and evolutionary approaches in dissecting molecular mechanisms in sexual size dimorphism	21
2.4.1	Whole genome bisulfite sequencing	22
2.4.2	Transcriptome sequencing	23
2.4.3	Estimation of evolution rates of coding genes	26
2.5	Overview of genome editing approaches to functionally characterize genes in teleosts	27
2.5.1	<i>In vivo</i> CRISPR/Cas9 mutagenesis in teleosts	29
2.5.2	<i>In vitro</i> CRISPR/Cas9 in teleost cell lines	30
2.6	Knowledge gap in understanding the molecular mechanisms of sexual size dimorphism in tilapia	31
2.7	Aims and objectives of this thesis	32
3	<i>Materials and methods</i>	38
3.1	Ethics statement	38
3.2	Fish husbandry	38
3.3	Traits measurement and visual sexing	39
3.4	Tissue collection	40
3.5	Molecular sexing with microsatellite markers	41
3.6	Whole genome bisulfite sequencing	42
3.6.1	DNA isolation from skeletal muscle for whole genome bisulfite sequencing	42
3.6.2	Bisulfite conversion of genomic DNA	43

3.6.3	Whole genome bisulfite sequencing (WGBS) library construction	44
3.6.4	WGBS bioinformatic analysis	46
3.7	Transcriptome analysis	48
3.7.1	RNA isolation from skeletal muscle and brain tissue for transcriptome analysis	48
3.7.2	mRNA sequencing library construction	49
3.7.3	Ribosomal depleted sequencing library construction	54
3.7.4	Transcriptome bioinformatics analysis	55
3.8	Candidate gene functional analysis <i>in vitro</i>	62
3.8.1	OmB cell culturing	62
3.8.2	CRISPR/Cas9 plasmid for in-vitro genome editing of OmB cells	63
3.8.3	Transfection of OmB cells with Lipofectamine 3000	64
3.8.4	DNA extraction from transfected cells and T7 Endonuclease Assay	65
3.8.5	Imaging of transfected cells	66
3.8.6	Prediction of estrogen response element in promoter region of POMC	66
3.8.7	Cloning of POMC ERE into Luciferase Vector	67
3.8.8	Transfection of pomcERE-pGL3-promoter into OmB cells and Luciferase Assay	68
3.8.9	Luciferase assay	69
3.9	Candidate gene functional analysis <i>in vivo</i>	69
3.9.1	Fish husbandry	69
3.9.2	Guide RNA design and <i>in vitro</i> synthesis	70
3.9.3	Microinjection of sgRNA and Cas9 protein	73
3.9.4	Genotyping of <i>pomc</i> mutants	73
3.9.5	Sequencing of genome editing sites	75
3.9.6	Phenotypic observation of <i>pomc</i> mutants	76
4	Results	79
4.1	Whole genome bisulfite sequencing	79
4.1.1	Single base-pair resolution methylome of tilapia	79
4.1.2	Tilapia gene body methylation patterns	82
4.1.3	Methylation profiles of various repeat elements in tilapia reference genome	83
4.1.4	Sexually dimorphic differentially methylated regions revealed by sliding windows analysis.	87
4.1.5	CpG methylation in promoters is negatively correlated with the gene expression	92
4.1.6	Whole genome bisulfite sequencing reveals sexually dimorphic regulatory landscape	93
4.2	Muscle transcriptome	98
4.2.1	Identification of sex-biased genes in tilapia muscle	98
4.2.2	Validation of RNA-seq data with qRT-PCR	99
4.2.3	Chromosomal locations of skeletal muscle sex-biased genes.	99
4.2.4	Identification of zebrafish orthologues using BLAST.	102
4.2.5	PantherGO and KEGG pathway mapping.	102
4.2.6	Female sex-biased genes exhibit higher evolution rates.	103
4.2.7	Identification of SNPs in RASGRF1 promoter region and its association with bodyweight in tilapia.	106

4.3	Brain Transcriptome	108
4.3.1	Sexing tilapia with microsatellite genotyping	108
4.3.2	RNA-seq read statistics and mapping efficiency	109
4.3.3	qPCR validation of RNA-seq data from brain tissues	111
4.3.4	Identifications of sex-biased gene expression in tilapia brain	111
4.3.5	Chromosomal locations of sex-biased genes	115
4.3.6	Gene ontology and KEGG pathways of sex-biased genes	118
4.3.7	Non-synonymous/Synonymous rate of mutations of sex-biased genes	128
4.3.8	Identification of SNPs from RNA-seq data	132
4.3.9	Association of SNP markers in chromosome LG1 with sex	132
4.4	<i>In vitro</i> genome editing of candidate genes in OmB cell lines	135
4.5	<i>In vivo</i> genome editing of candidate gene in zebrafish	136
4.5.1	<i>POMC</i> in Mozambique tilapia	136
4.5.2	Possible functional estrogen receptor elements upstream of <i>POMC</i> transcriptional start site	138
4.5.3	Establishment of a <i>POMC</i> -knockout zebrafish line as a model to study somatic growth and feeding behavior	140
4.5.4	F0 mutants showed smaller melanophores coverage	146
4.5.5	Increased growth observed in F1 heterozygous <i>pomc</i> zebrafish	150
4.5.6	<i>pomc</i> juveniles showed shorter feeding reaction to feeding in comparison to wildtype	152
5	<i>Discussion and future works</i>	154
5.1	Whole genome bisulfite sequencing	154
5.1.1	Advantage of single base resolution whole genome bisulfite sequencing over other methodology	155
5.1.2	Epigenetic architecture of tilapia is similar to other vertebrate animals	156
5.1.3	Repeat elements in tilapia genome are highly methylated	157
5.1.4	CpG methylation in promoter regions is correlated with gene expression level in tilapia	158
5.1.5	Tilapia methylome is sexually dimorphic in somatic tissue	159
5.1.6	Inference of sexually dimorphic active cis-Regulatory Elements	160
5.1.7	Future works in tilapia CpG methylation analysis	161
5.2	Muscle transcriptomes	162
5.2.1	Sex-biased gene expression in tilapia skeletal muscle	162
5.2.2	Pathways regulating cell growth and proliferations are sexually biased in expression	163
5.2.3	Sex-biased genes located in QTLs related to growth	165
5.3	Brain transcriptome	166
5.3.1	Use of microsatellite markers for molecular sexing	166
5.3.2	Sex biased gene expression in the brain	167
5.3.3	Sex-biased genes in the brain are not randomly distributed in the genome	169
5.3.4	Hypothalamic-Pituitary Axes are overrepresented with sex-biased genes	170
5.4	<i>POMC</i> as a candidate gene for sexual size dimorphism in tilapia.	172
5.4.1	<i>pomc</i> mutants have smaller melanophore distribution compared to wildtype	174
5.4.2	<i>pomc</i> mutants exhibit higher growth rate	175

5.4.3	<i>pomc</i> mutants exhibit short feeding latency behaviour in comparison with wildtype fishes _____	176
5.4.4	Transcription of <i>POMC</i> in tilapia may be induced by estrogen _____	177
5.4.5	Proposed model of <i>POMC</i> mechanism in tilapia sexual size dimorphism ____	179
5.4.6	Future works in <i>POMC</i> and its possible applications in aquaculture _____	181
5.5	Single nucleotide polymorphism markers for marker-assisted selection.	183
5.6	Sex-biased genes in tilapia are under relaxed purifying selection. ____	185
6	<i>Concluding remarks.</i> _____	186
7	<i>References</i> _____	189

1 **1 Abstract**

2

3 Tilapia is an important aquaculture species. Males outgrow females. I explored
4 the genetic and epigenetic architecture of male and female tilapia to understand
5 more about sexual dimorphism. Whole genome bisulfite sequencing and RNA-
6 seq in skeletal muscle revealed sexually-dimorphic methylated regions and sex-
7 biased gene expression. *dN/dS* analysis uncovered that sex-biased genes in tilapia
8 somatic tissues were under relaxed purifying selection. Analysis of brain
9 transcriptomes identified 124, 55 and 2706 sex-biased genes at 5, 30 and 90 days
10 post hatch, respectively. The pro-opiomelanocortin (*pomc*) gene, which was
11 female-biased in the brain, was selected for functional analysis. *pomc* knock-out
12 zebrafish showed faster growth and higher sensitivity to feeding compared to
13 wildtype. Two estrogenic response elements upstream of tilapia *pomc* were
14 sensitive to estrogen induction in a luciferase reporter assay. These results
15 suggest that *pomc* may be involved in sexual size dimorphism. In addition, I
16 identified SNPs located upstream of *rasgrfl*, which were associated with
17 increased growth rate and may be useful in selecting fast-growing tilapia.

18

19 2 Introduction

20

21 2.1 Sexual dimorphism and its prevalence in teleosts

22 Sexual dimorphism refers to the significant anatomical, physiological and
23 behavioural differences between male and female individuals in a given species¹.

24 They exist at various time points throughout the life spans of organisms²⁻⁶. These
25 phenotypic differences may confer genetic fitness in traits that promote the
26 survivability of a species should the traits be subjected to evolutionary
27 selections^{2,7,8}. Despite the widespread prevalence of sexually dimorphic traits
28 across the animal kingdom, knowledge on the origin, mechanisms and evolutions
29 of sexually dimorphic traits is still in its infancy. Understanding how these traits
30 developed differently in two sexes despite the constraint of sharing the same
31 genome is essential for identifying causative genomic regions for sexual
32 dimorphism as well as the molecular pathways leading up to sexual
33 dimorphism^{3,4,9}.

34 Sexually dimorphic traits can be classified into primary sex traits and
35 secondary sex traits¹⁰. Primary sex traits refer to the anatomical and physiological
36 traits that contribute directly to sexual reproduction. Examples for primary sex
37 trait categories are gonads and various sexual organs that enable organisms of
38 two different sexes to copulate and reproduce sexually. Secondary sex traits are
39 modifications to aesthetics, anatomy and behaviours that benefit an individual's
40 reproductive success. Examples of traits that fall under this category are colourful
41 male guppy tails, long elaborate horns of male rhinoceros beetles and big noses

42 of male proboscis monkeys¹⁰. These physical traits are often ornamental traits
43 that enhance the reproductive success of a male.

44 In teleost fish, sexual dimorphism takes form in various aspects including
45 colour, size, morphology and behaviors. One of the most striking sexually
46 dimorphic phenotypes in teleosts is sexual size dimorphism². In fish, sexual size
47 dimorphism can be observed to be male-biased, female-biased, or in some species,
48 no differences can be observed². In male-biased species, such as Nile tilapia
49 (*Oreochromis niloticus*) and Mozambique tilapia (*Oreochromis mossambicus*),
50 males outgrow females by up to 40% in certain strains^{11,12}. An extreme male-
51 biased sexual size dimorphism was observed in the African cichlid, *Lamprologus*
52 *callipterus*, where males grew up to 12 times the weight of females¹³. In wild
53 salmonid species, male-biased sexual size dimorphism can also be observed in
54 anadromous breeding adults^{5,14,15}.

55 Female-biased sexual size dimorphism can also be observed in teleosts. In
56 half-smooth tongue sole (*Cynoglossus semilaevis*), a female-biased sexual size
57 dimorphism is observed, with each female achieving body-size up to 4 times of
58 that in a male^{6,16}. Another species that exhibits female-biased sexual size
59 dimorphism is the Atlantic Cod (*Gadus morhua*), whereby females are on
60 average 10% larger than males¹⁷. It has been shown that larger female size in *G.*
61 *morhua* confers reproduction advantage as larger female cod reproduce
62 disproportionately more than smaller female cod¹⁷.

63 The phenomenon of sexual size dimorphism in teleost across many genera
64 is well studied in terms of size and allometry. Evolutionary theories have been
65 proposed on how sexual size dimorphism is under selection to confer genetic

66 fitness. It has been proposed that female-biased sexual size dimorphism may be
67 favoured for natural selection as fecundity and reproductive success is higher
68 with larger body size². This theory is particularly attractive for fishes that
69 reproduce through mass mating or mate in a polygynandry manner, as larger egg
70 numbers are produced with increasing body size. However, this selective
71 pressure is balanced out by the mortality of female fish from predator or disease
72 before growing out to achieve sexually matured size.

73 For male-biased sexual size dimorphism, the model proposed is that larger
74 males may be favoured through both natural selection and sexual selection². In
75 the natural selection model, larger males are at the advantage of producing more
76 milt with increasing body size, hence allowing a higher rate of fertilization
77 success in a polygynandrous breeding set². In the sexual selection model, larger
78 males with faster growth rate may have the advantage in male-to-male combat
79 during breeding seasons and be favoured by females on the breeding grounds^{11,18}.
80 Under a no environmental constraint model, larger males may have access to
81 more mates and allow for a higher number of spawning than smaller males². On
82 the contrary, male size can be limited by predation and/or mortality before sexual
83 maturity. Hence, in some species occupying certain ecological niche, achieving
84 sexual maturity early may be an advantage.

85

86 2.2 Biology of tilapia

87 2.2.1 Overview of tilapia biology and tilapia as an aquaculture species

88 Originating from the continent of Africa, Tilapia is one of the most economically
89 important teleosts that exhibits sexual dimorphism^{11,12}. The common name,
90 Tilapia, actually refers to close to a hundred species of Cichlid fishes under the
91 Tilapiini tribe within the larger order of Perciformes¹⁹. Tilapiini includes three
92 different genera including *Tilapia*, *Sarotherodon* and *Oreochromis*, which are
93 grouped based on their spawning habits including substrate spawner (*Tilapia*),
94 paternal mouthbrooder (*Sarotherodon*) or maternal mouthbrooder
95 (*Oreochromis*)¹⁹. Amongst these genera, the *Oreochromis* genus contains some
96 of the most important aquaculture species, with an estimated combined
97 production of 5,377,000 tonnes in 2018²⁰. These species include the Nile tilapia
98 (*Oreochromis niloticus*), Mozambique tilapia (*Oreochromis mossambicus*) and
99 *Oreochromis aureus*. The most cultured species is *O. niloticus*, with a global
100 production output of 4.2 million tonnes in 2018²⁰. This is followed by *O.*
101 *mossambicus*, which has a reported aquaculture output of 38.5 thousand tonnes
102 in 2018²¹. *O. niloticus* originates from the inland freshwater riverine regions of
103 central Africa and the Nile River, which its common name is derived from¹⁸. *O.*
104 *mossambicus* is a native of coastal regions on Eastern Africa such as the Zambezi
105 river and its surrounding brackish waters along the coastlines of Mozambique¹⁸.
106 This species can survive in freshwater, brackish water and hypersaline water,
107 making it an ideal candidate for saltwater tilapia aquaculture²².

108 Tilapia has excellent characteristics which make it an ideal candidate for
109 aquaculture¹⁸. Also known as the “aquatic chicken”, tilapia is highly adaptable to
110 diverse habitats and has high feed conversion efficiency while displaying a
111 relatively high rate of growth compared to other species¹⁸. Tilapia is also one of
112 a few aquaculture species with several well documented selective breeding
113 programs, with the most successful program being the GIFT (Genetically
114 Improved Farmed Tilapia), with growth improvement up to 10% per generation
115 reported²³.

116 2.2.2 Sexual dimorphism in tilapia and its implication to aquaculture

117 *Oreochromis* tilapia exhibit several sexually dimorphic traits, including larger
118 mouth, bigger dorsal and anal fins, as well as faster growth rate in males^{11,24}.
119 These are traits that are important for agonistic display (larger body size, strong
120 mouth for male to male combat and large fins for territorial display)²⁴. The trait
121 of focus in this thesis will be male and female tilapia body weight differences
122 post-sexual maturation¹¹.

123
124 Figure 1. Sexual dimorphism between female (left) and male (right) Mozambique
125 tilapia (*Oreochromis mossambicus*) includes larger mouth, higher dorsal fins,
126 and greater size in males. Tilapia pair shown in picture are 120 days post hatch.
127

128 In *O. niloticus*, *O. mossambicus*, *O. aureus* and *O. shiranus*, males exhibit
129 a faster growth rate than females in natural habitat and in aquaculture settings¹².

130 Female to male body weight ratio at sexually matured state was reported to be
131 between 52.2% to 84.7% (Female:Male bodyweight)¹². It was reported that
132 different strains of tilapia show different ratios of sexual size dimorphism¹².
133 However, as these studies were conducted with fishes raised in different locations,
134 it is not possible to pinpoint if these variations are due to the environment or
135 genetic factors. Besides environmental and genetic effects, the effects of social
136 interactions may also influence the degree of sexual size dimorphism in tilapia.
137 *O. niloticus* females have been reported to grow significantly larger than their
138 male counterparts when raised individually instead of mixed-sex or mono sex
139 culture¹². It was hypothesized that the slower growth rates in females may be
140 caused by behavioural factors during a mixed-sex culture rather than
141 physiological factors¹².

142 Due to these differences in yield sizes of different sex between male and
143 female tilapia, production of monosex tilapia stock for aquaculture has been
144 favoured by fish farmers as a means to improve production²⁵⁻³⁰. To produce
145 monosex fries, 5 days post hatch (dph) fries were fed with 17 α -
146 methyltestosterone supplemented feed (60 mg/kg feed) until 30 dph^{25,26,28,30}. This
147 would lead to masculinized all-male monosex tilapia stocks. However, this
148 technique of producing all-male tilapia fries came under the scrutiny of critics
149 regarding environmental and health concerns³¹. Certain consumer markets, such
150 as the European Union, banned the importation and production of monosex
151 tilapia using hormone treatments under Directive 2003/74/EC³¹.

152 An alternative to hormone masculinization of tilapia is by using YY-
153 supermale brood stocks^{26,32}. Offspring generated from crossing XX female and

YY supermale are, in theory, all XY males, assuming no environmental interactions such as temperature which may skew the sex ratio. To create a YY supermale brood stock, XY males are feminized with feed supplemented with 17 α -ethinylestradiol from 5 dph until 30 dph³². Some XY pseudo-males are capable of developing mature oocytes and were used to mate with normal XY males to generate up to 25% of YY males of the entire cohort. This method is considered to be more environmentally friendly compared to the hormone masculinization method described previously. However, technical challenge remains in distinguishing YY males from XY males reliably as well as in generating fertile XY pseudo-males²⁵. In many cases, feminized XY males were not fertile and have an inter-sex gonad, leading to difficulties in generating, large scale, XY females for YY tilapia production. A basic understanding of the male-determining allele in tilapia is also necessary to enable sex genotyping for distinguishing YY males from XY males, as both fish are phenotypically male.

2.2.3 Sex-determining loci in tilapia

Due to the importance of sex control and the economic benefits of an all-male cohort in tilapia aquaculture, many research groups have attempted to identify the sex-determining gene in various tilapia species. In *O. niloticus*, the sex-determining locus was first identified to be in linkage group 8, using microsatellite marker and bulked segregation analysis³³. Subsequent efforts to map loci associated with sex determination, sex-specific mortality or both, identified linkage groups 1, 2, 3, 6 and 23 to be associated with sex³⁴. Then, subsequent fine mapping approaches narrowed down the candidates to only linkage group 23^{34,35}. Using SNP markers genotyped with double digest

178 restriction-site associated DNA sequencing (ddRADseq) approach, sex-
179 determining loci were mapped to linkage groups 1, 3 and 23 in a multi-family
180 quantitative trait locus (QTL) study³⁶. With the emergence of CRISPR/Cas9
181 technology, the Y-specific duplicate of anti-Müllerian hormone (*AMHy*), a
182 candidate gene located in the sex-determining locus in linkage group 23, has been
183 successfully knocked out³⁷. Loss-of-function in *AMHy* resulted in complete
184 male-to-female sex reversal, while overexpression of *AMHy* in XX transgenic
185 females resulted in female to male reversal³⁷. Taken together, these results
186 suggest that *AMHy* located on chromosome LG23 is a major sex-determining
187 gene in *O. niloticus*.

188 In *O. mossambicus*, the sex-determining locus was found to be on linkage
189 group 1, identified using a fluorescence in situ hybridization (FISH) assay³⁸.
190 Subsequent QTL analysis confirmed these findings using microsatellite
191 markers^{39,40}. The sex-determining locus in linkage group 1 was also found to be
192 associated with growth, explaining up to 71.7% of phenotypic variance in body
193 weight⁴⁰. Several candidate genes were located in this region, including estrogen-
194 related receptor alpha (*esrra*), growth regulation by estrogen in breast cancer 1
195 (*greb1*), spermatogenesis associated 17 (*spata17*) and zinc finger AN1-Type
196 containing 3 (*zfang3*). Although the candidate sex-determining gene was not
197 identified here, these findings provided microsatellite markers to be used to
198 identify sex in *O. mossambicus*, for example, at juvenile fry stage⁴⁰. However,
199 microsatellite markers tend to be family-specific, hence pedigree-specific allele
200 information is required to trace the sex-determining locus.

201
 202 Figure 2. Linkage map of chromosome LG1 in Mozambique tilapia showing the
 203 marker S1543 which is linked with sex phenotype as shown by Liu et al⁴⁰. The
 204 location of each marker is indicated on the left side of the linkage group in
 205 Kosambi centimorgans.

206 2.2.4 Availability of genetic and genomic toolkits in tilapia

207 In tilapia, several types of genetic resources have been developed. Even before
 208 the completion of the tilapia reference genome, microsatellite markers and SNP
 209 markers had been developed and used in multiple applications including
 210 phylogeny assignments⁴¹, parental assignments⁴², genetic conservation^{22,43},
 211 development of genetic linkage map^{34,35,39}, QTL analysis^{36,40,44} and for detecting
 212 signatures of selections⁴⁵. As tilapia is an economically important food fish, the
 213 majority of the markers developed were used for QTL mapping and analysis of
 214 economically important traits such as growth rates^{40,44}, sex determination^{36,44} and
 215 cold-resistance⁴⁶. SNP-arrays containing up to 58K SNPs⁴⁷ and ddRAD-seq⁴⁸

technologies were also developed, enabling high throughput marker screening and high-density linkage maps to be constructed for high throughput analyses such as genome-wide association studies (GWAS) and marker-assisted selection of economically important traits. Besides linkage maps, a well-assembled and annotated reference genome is also critical for understanding the genetic architecture of complex traits such as sexual dimorphism in tilapia. As of the writing of this thesis, four versions of *O. niloticus* reference genomes have been published^{49,50}.

2.3 Current understanding of sexual size dimorphism in teleosts

Physiological and molecular biology research regarding sexual size dimorphism has been conducted by several groups but are still at its inception stage. Several techniques including gonadectomy⁵¹, hormone treatment^{26,52}, reverse genetic approach⁵³ and transcriptomic analyses^{6,16} have been employed to dissect the mechanisms of sexual size dimorphism in various teleost species. Many of these species are important in aquaculture and sexual size dimorphism has a direct impact on their aquaculture yield and profitability. Here, major new findings with regards to sexual size dimorphism in teleost and other model organisms are highlighted. These include the role of gonads in regulating somatic growth in fish⁵⁴, pathways involved in growth regulations in fish^{12,16,40,44,53,55-57}, the role of sex-biased gene expression in somatic tissues^{16,58-60}, differential DNA CpG methylation in controlling gene expression⁶¹⁻⁶⁵ and evolutionary aspect of sexual dimorphism with regards to sexual conflict^{58,66-70}.

239 2.3.1 Teleost growth regulation mechanisms

240 Teleost growth has been reported to be regulated by the pituitary gland⁷¹⁻⁷⁴.
241 Hormones produced from this gland include growth hormone (Gh), growth
242 hormone-releasing hormone (Ghrh) and somatostatin (Srif). Besides growth and
243 development⁷⁵, these hormones also regulate osmoregulation^{76,77}, immune
244 function^{78,79}, metabolism⁷⁹ and feeding behaviour⁸⁰. Teleost growth rate reflects
245 the balancing between anabolism and catabolism, which is the process of
246 building up and breaking down molecules and cells respectively^{72,73,81}. Biological
247 processes such as reproduction^{75,82} and food availability⁸³ will impact the growth
248 rate of a teleost, as energy homeostasis needs to be diverted for these processes.

249 The molecular pathways regulating the rate of growth in teleosts are well
250 studied, especially in terms of endocrinology. Teleost pituitary glands release
251 growth hormones, which will bind directly to growth hormone receptors on
252 skeletal muscle sarcolemma and in liver cells. Binding of Gh to its receptors on
253 liver cells induces the secretion of insulin-like growth factor 1 (Igf1) from the
254 liver^{52,76,77,79}. Igf1 binds to insulin-like growth factor receptor 1 (Igf1r) located
255 on cell membranes of target cells, leading to the auto-phosphorylation of
256 Igf1r^{73,84}. This will then activate several downstream signal transduction
257 cascades through signalling peptides such as insulin receptor substrate 1
258 (*irs1*)^{85,86}. One of the pathways induced by this signalling cascade is the *PI3K*-
259 *Akt-mTOR* pathway, leading to activation of *mTOR* via phosphorylation by *Akt*⁸⁷.
260 Activated mTOR recruits five other protein complexes to form *mTOR* complex
261 1 (*mTORC1*), which is the master regulator of cell growth and metabolism^{87,88}.
262 *mTORC1* promotes protein synthesis via inhibition of initiation factor 4E

263 (*eIF4E*)-binding protein 1 (*4E-BP1*) to promote cap-dependent protein
264 translation⁸⁸. Besides promoting protein synthesis, *mTORC1* also activates
265 *LPIN1* by phosphorylation and induces lipid synthesis, which is another hallmark
266 of anabolic growth⁸⁸. Besides *IGF-1* and growth hormone, other growth factor
267 signalling pathways may also promote cell growth and proliferations. The *Ras-*
268 *MEK-ERK* pathway is an example of how cells can accept external growth factors
269 and trigger a cascade of phosphorylation in the cell cytoplasm and activation of
270 transcription factors in the cell nucleus which leads to cell proliferation^{89,90}.

271 Another emerging pathway which have been shown to increase somatic
272 growth is the melanocortin system. The melanocortin system regulates several
273 physiological process including complexion, steroidogenesis, energy
274 homeostasis, exocrine secretion, emperature control, cardiovascular regulation,
275 and neuromuscular regeneration. One possible way for the melanocortin system
276 to regulate somatic growth is via the control of feed intake and energy
277 homeostasis. This system operates via a push-and-pull mechanism via two
278 component genes, namely *pomc* and *agrp*. Within the arcuate nucleus of the
279 vertebrate brain, two sub-populations of neurons expresses these genes, namely
280 the *POMC* neurons and *AgRP* neurons. *POMC* neurons secretes *Pomc* to
281 surrounding cells with melanocortin receptors to exert its function of inhibiting
282 food intake. *AgRP* neurons on the other hand, promotes feeding behaviour. These
283 two hormones work in tandem to regulates the feeding behaviour in animals,
284 hence providing a regulatory machinery on nutrition availability for promoting
285 somatic growth.

286 **2.3.2 Teleost gonads directly regulate somatic growth**

287 Besides the hypothalamic-pituitary axis, another unique feature in teleosts is that
288 gonads play essential roles in regulating growth. In *O. mossambicus*, it has been
289 shown that gonads directly regulate growth⁵¹. Gonad removal experiments in
290 both male and female Mozambique tilapia stunted growth in both sexes when
291 compared to sham surgery fishes (incision on the abdominal cavity without
292 removal of gonads)⁵¹. Using real time PCR, the authors discovered that gonads
293 do express growth hormones and that gonadal growth hormone has a significant
294 contribution towards somatic growth in both males and females⁵¹. Gonadal
295 transplantation from the original position at the abdominal cavity into the dorsal
296 skeletal muscle regions can rescue the stunted growth phenotype in
297 gonadectomized Mozambique tilapia⁵¹.

298 Besides Mozambique tilapia, Japanese eel (*Anguilla japonica*) also
299 exhibits expression of growth hormone, growth hormone receptor 1 and growth
300 hormone receptor 2 in the testis of male eels⁵⁴. A study on Japanese eel
301 spermatogonia proliferation identifies a role for eel growth hormone in the
302 regulation of early spermatogenesis, particularly in the mitotic phase of
303 spermatogenesis that is not mediated by either steroid hormones or insulin-like
304 growth factor 1 production⁵⁴. These reports showed that gonads play important
305 roles in regulating somatic growth in teleosts, through the secretion of growth
306 hormones and/or other endocrine factors. However, one major open question
307 remains on how gonads or other factors contribute towards sexual size
308 dimorphism in teleost? One possible route is via the sex-biased gene expressions
309 in somatic tissues.

310

311 2.3.3 Sex-biased gene expression in teleost somatic tissue

312 Besides the roles played by the expression of growth hormones in the gonadal
313 tissue, sex-biased gene expression in somatic tissues may have implications in
314 causing sexual size dimorphism in fish. With the development of next-generation
315 sequencing technology and advanced de-novo assembly algorithms for
316 transcriptomes such as Trinity⁹¹, transcriptomic studies have been liberalised for
317 non-model organisms to be conducted on a genome-wide scale. These studies
318 provided exploratory insights into the genes that are sex-biased in expression,
319 although the expression level of these genes varies spatially and temporally at
320 different tissues and developmental stages. Sex-chromosome specific processes
321 such as X-chromosomal inactivation⁹², dosage compensation⁹³ or trans-
322 regulatory elements originating from sex-chromosomes^{94,95} may also influence
323 sex-biased gene expression.

324 In *Cynoglossus semilaevis*, which is a species that shows female-biased
325 sexual size dimorphism, transcriptomic profiling of the brain, pituitary, liver, and
326 gonad from male and female adults highlighted several molecular pathways,
327 which may contribute towards sexual size dimorphism in *C. semilaevis*^{6,16}. These
328 pathways include nucleic acid-binding transcription factor activity, G-protein-
329 coupled receptor activity, *mapk* signalling pathway, steroid biosynthesis, and
330 neuroactive ligand-receptor interaction. These pathways have also been reported
331 to be related to somatic growth in various species^{53,73,77,88,89}.

332 Another species, the guppy, *Poecilia reticulata*, also exhibits strong
333 sexual dimorphism in size, ornaments and behaviours⁵⁸. In sexually dimorphic

334 guppy, tissue-associated sex-biased expression of genes was found to be related
335 to pigmentation, signal transduction, and spermatogenesis in males; and growth,
336 cell-division, extra-cellular matrix organization, nutrient transport, and
337 folliculogenesis in females^{58,70}. Accumulation of genes related to ovarian
338 function on the putative X-chromosome was also found in this species⁵⁸. This
339 supports the hypothesis that genes advantageous to only one sex will accumulate
340 on the differentiating sex chromosomes⁵⁸. The rapid evolution of testis-biased
341 genes and female-biased genes expressed in the brain and the liver were detected
342 based on elevated dN/dS values⁵⁸. Sex-biased coding genes have been shown to
343 be under divergence, in terms of the coding sequence and expression patterns,
344 between closely related species^{96,97}. These genes showed elevated non-
345 synonymous to synonymous substitution rates (dN/dS) which suggests an
346 accelerated evolution rate due to positive or sexual selections. Such phenomena
347 could be the effect of several events such as sexual selection, sexual antagonism,
348 and relaxed selective constraint. In fish, this has been shown to be true for both
349 male and female-biased genes⁹⁸. In animal models such as *Drosophila*, only
350 male-biased genes have been shown to be under selection⁹⁹. To add to the
351 complexity of these observations, the majority of these genes are different in
352 biological function and cellular locations⁵⁸. Hence, it is possible that males' and
353 females' sex-biased genes are under different selection pressures, independent of
354 the breadth of tissue expression⁵⁸.

355 Sex hormones may also trigger sexual dimorphism by inducing sex-
356 biased gene expressions. One example was reported in medaka (*Oryzias*
357 *latipes*)⁵⁶, which shows a prominent masculine sexual characteristic for

358 appendage development which is the formation of papillary processes in the anal
359 fin. This formation of papillary processes can develop in females by exogenous
360 androgen exposure. It was shown that the androgen receptor, when driven by
361 androgen in the form of methyltestosterone, can promote the expression of bone
362 morphogenic protein 7 (*bmp7*) and lymphoid enhancer-binding factor-1 (*lef1*).
363 Androgen driven expressions of *bmp7* and *lef1* are essential for the bone
364 outgrowth which leads to the formation of anal fin secondary sex characteristics
365 in medaka. This provided examples that androgen produced by the testis have
366 direct implications on somatic organ sex-biased expression and differential
367 development, which leads to sexual dimorphism.

368

369

370 **2.3.4 Differential CpG DNA methylation between sexes**

371 Besides genetic factors, phenotypes are also determined by environmental factors.
372 Gene regulation involves modulating the spatial and temporal modulation of gene
373 transcription¹⁰⁰. One method of suppressing gene expression is via DNA CpG
374 methylation¹⁰¹. The addition of a methyl group to the fifth carbon of a cytosine
375 base (5mC) and adenine methylation are the major forms of DNA methylation,
376 with CpG methylation being the most studied form. Other forms of cytosine
377 methylations are CHH and CHG, with H representing adenine (A), guanine (G)
378 or thymine (T). These stable epigenetic markers are heavily involved in many
379 cellular functions including cell differentiation, gene regulation, suppressing
380 transposable elements, germ cell formation, genomic imprinting and X-

381 chromosome inactivation¹⁰². There is evidence that gene regulation has a major
382 role in creating sex-specific differences. For example, in *Drosophila*
383 *melanogaster*, sexually dimorphic abdominal pigmentation was found to be due
384 to differences in regulation between males and females. One example of gene
385 regulation is DNA methylation on CpG islands located in the promoter region of
386 the targeted gene. Recent studies showed that epigenetic factors such as DNA
387 methylation may play an important role in sexual dimorphism⁶². Recent
388 developments in CpG methylation analysis have shed light on the function and
389 information content of DNA methylation in terms of cis-regulatory functions¹⁰³⁻
390 ¹⁰⁵. Reports have shown that DNA methylation marks and transcription factor
391 binding motifs can serve as a highly informative epigenetic marker to map cis-
392 regulatory elements in a given genome^{103,106}. These cis-regulatory elements
393 include genomic elements such as enhancers and promoters which are
394 responsible for regulating development and physiology by acting as a control
395 switch for gene expression^{103,106}.

396 Current understanding of fish methylome is still in its infancy. Most
397 research on fish methylome is focused on fish gonadal tissue to investigate the
398 effects of hormonal and temperature treatment on sex-reversal in teleosts^{62,64}.
399 Very few datasets are currently available on investigating the effect of gender on
400 the DNA methylation profiles on somatic tissues.

401 **2.3.5 Molecular evolution of sexual dimorphism**

402 Sexual dimorphism may have been the consequence of natural and/or sexual
403 selection acting on traits that give rise to differences between sexes in phenotypes

404 and behavioural differences^{4,93,96,97,107}. Different selective pressures on these
405 traits may contribute to differences in transcriptional differences or translational
406 differences. In some cases, selection pressure may even lead to a gene being
407 silenced in one sex but actively expressed in another sex^{96,108}. These expression
408 differences of sex-biased genes may be the result of either natural selection on
409 reproductive potential and survival traits of either sex or sexual selection on traits
410 relevant to intra-sexual competition and inter-sexual mate choice^{4,96,97}. These
411 hypotheses form the core idea on the evolutionary origin of sex-biased genes, as
412 sex-biased genes may confer reproductive success in one sex but at the cost of
413 survivability in another sex⁷⁰.

414 Besides sexual selection and natural selection which may be used to
415 explain the evolutionary origin of sex-biased genes that may lead to sexual
416 dimorphism, a third phenomenon may also explain sex-biased gene expression –
417 sexual antagonism, also known as sexual conflict in some literature^{69,70,99,109,110}.
418 Sexual antagonisms occur when selection acts in opposite directions on males
419 and females¹⁰⁹. It represents an evolutionary conflict acting on one or multiple
420 loci that contribute differentially to male and female fitness. As some alleles are
421 beneficial for one sex but reduce the fitness of another sex, counter adaptations
422 such as suppressor alleles can develop at different genetic loci to reduce the
423 effects of the afore-mentioned allele, thus giving rise to differences in gene
424 expressions^{107,108}. Hence, sex-biased expression can be interpreted as a
425 mechanism evolved to resolve sexual conflict in a system^{58,93,111-116}. There are
426 two types of sexual conflict commonly studied in evolutionary biology, namely
427 inter-locus sexual conflict and intra-locus sexual conflict.

428 Inter-locus sexual conflict is the evolutionary antagonistic interactions
429 between males and females for alleles at two or more loci¹¹⁰. In inter-locus sexual
430 conflict, a new allele that favours one sex but incurs a cost for the other sex is
431 counter adapted to another allele at a different locus¹¹⁷. These cycles of
432 adaptation and counter-adaptation can manifest many times in a cyclical manner,
433 resulting in an arms race between males and females. Although this evolutionary
434 arms race is an interesting biological question, research investigating inter-locus
435 sexual conflict is somewhat limited as it is difficult to study due to the lack of
436 clear evolutionary signature in the DNA sequence.

437 Intra-locus sexual conflict refers to the conflicting selection pressures for
438 males and females over alleles at a single locus. In this mode of sexual
439 antagonism, the alleles on a single locus have different effects on male and female
440 fitness. Similar genetic architecture underpinning both the sexes in a given
441 species will give rise to intra-locus sexual conflict. Besides the allosomes (Y and
442 W chromosomes), males and females share almost the same genome. Studies
443 involving reproductive fitness in *D. melanogaster* have shown that intra-locus
444 sexual antagonisms happen at multiple loci across the genome⁹⁶. One example of
445 intra-locus sexual conflict is the locomotory traits in *D. melanogaster*. A sexually
446 antagonistic selection on locomotory activity were detected in a laboratory-
447 adapted population. Briefly, males with the highest locomotory activity showed
448 higher courtship and reproductive success. Mutations which affects locomotor
449 activities in males reduced reproductive success in males. Cumulatively, these
450 observations suggests a sexual selection for increased levels of locomotory
451 activity in male *D. melanogaster*. However, in females, decreased fitness is

452 associated with increased locomotory activities. Since female movement
453 activities are stimulus for male courtship, increased locomotory in females leads
454 harmful persistent male courtship. Also, persistent courtship in female will cause
455 reduced life-time offsprings in comparison with females with less courtship
456 activity. Hence, the conflicting outcome of having higher locomotry traits in *D.*
457 *melanogaster* is one of the example of intra-locus sexual conflicts. These intra-
458 locus sexual antagonisms are then resolved through allele-specific expression or
459 sex-specific expression patterns, which includes sex-biased gene expression.

460

461 **2.4 Overview of genomics and evolutionary approaches in dissecting** 462 **molecular mechanisms in sexual size dimorphism**

463

464 Conventionally, genetic analysis relies heavily on a candidate gene approach to
465 understand the biological question on the molecular pathways involved in sexual
466 dimorphism. However, the candidate gene approach is laborious, costly and
467 inefficient. Furthermore, the lack of genetic toolkits in non-model organisms
468 made molecular genetic studies even more difficult. Sexual size dimorphism is a
469 complex trait controlled by multiple loci and multiple organs at different
470 developmental stages^{2,8,10}. Hence, adopting a candidate gene approach initially
471 to investigate the mechanisms behind sexual size dimorphism may not be
472 efficient.

473 Recent developments on massive parallel short sequencing technology or
474 “Next-Generation Sequencing” (NGS) technologies have provided researchers
475 with the technology required to understand complex biological questions such as

476 sexual size dimorphism¹¹⁸⁻¹²⁰. The high-throughput nature of NGS technology
477 allows investigation on genome-wide gene expression within short hands-on time
478 on laboratory works. In this section, I will discuss the two main applications of
479 NGS for dissecting the molecular mechanisms of sexual size dimorphism in
480 tilapia.

481 2.4.1 Whole genome bisulfite sequencing

482 The discovery that treatment of denatured genomic DNA with sodium bisulfite
483 chemically deaminates unmethylated cytosine residues much more rapidly than
484 methylated cytosines, spurred a revolution in DNA methylation analysis in the
485 1990s^{101,102}. This chemical treatment of DNA effectively turns an epigenetic
486 difference into a genetic difference — unmethylated Cs are converted to Ts (by
487 uracil) — thereby enabling many new DNA methylation detection and analysis
488 possibilities^{121,122}. Analysis of bisulfite-converted DNA was initially done by
489 Sanger sequencing of cloned PCR products from single loci¹²³. Many
490 enhancements have since been developed, including the direct Sanger sequencing
491 of PCR products and the highly automated application of this approach using
492 parallel Sanger sequencer such as ABI 3730xl¹¹⁹. However, the prohibitive costs
493 of Sanger sequencing have limited this technique to small subsets of loci only.

494 With the introduction of next generation sequencing such as Illumina's
495 sequencing-by-synthesis approach, many researchers have turned to NGS-based
496 methods for bisulfite sequencing¹²¹. Using standard bisulfite conversion
497 techniques, DNA templates containing unmethylated CpG sites are converted
498 into single-stranded uracil-containing template¹²¹. Unmethylated bacterial
499 Lambda DNA is spiked into each DNA sample as an internal control prior to

500 bisulfite conversion for measuring bisulfite conversion efficiency¹²¹. These
501 templates are then subjected to PCR amplifications and used as a standard DNA
502 template for NGS library construction. However, whole-genome bisulfite
503 sequencing can be costly, even at 10x sequencing coverage as the experiment
504 requires the entire genome to be sequenced at least 10 times. This approach may
505 be cost-prohibitive for studies involving population scale samples¹²⁴. An
506 alternative approach will be reduced representation bisulfite sequencing, which
507 incorporates restriction enzyme digestion and bisulfite conversion to enrich for
508 DNA fragments containing CpG methylation markers^{119,124}. This reduces the
509 amount of sequencing coverage needed for bisulfite sequencing project, hence
510 lowering cost. The development of computational tools and resources for DNA
511 methylation analysis is accelerating rapidly. Sequence-based analyses involve
512 alignment to a reference genome, quantification of DNA methylation and
513 differential methylated region analysis¹²¹. Dissecting these data onto regions of
514 different genomic elements such as repeat elements, promoters, exon, introns and
515 UTRs also can provide valuable insights into the biological questions
516 inspected¹²¹.

517 **2.4.2 Transcriptome sequencing**

518 The development of novel high-throughput DNA sequencing methods has
519 provided a new method for both mapping and quantifying transcriptomes. This
520 method, termed RNA-seq (RNA sequencing), has many advantages over
521 capillary based Sanger sequencing approaches and is revolutionizing the manner
522 in which eukaryotic transcriptomes and gene expressions are analysed^{120,125}.
523 RNA-seq takes advantage of the massively parallel, short read lengths and high-

throughput nature of NGS machines to collect up to millions of reads of expressed RNA per sample, providing a qualitative and quantitative data output¹²⁰. In general, a population of RNA (total or reduced representation such as poly A-tailed mRNA) is converted to a library of cDNA fragments with adaptors attached to both ends¹¹⁸. Each cDNA fragment is then anchored onto a flow cell embedded with oligo complementary to the adaptor sequences of the cDNA library. Through a process known as “bridge amplification”, a cluster of identical cDNA library is generated on a spot within the flow cell. Each cluster is then sequenced in a high-throughput manner to obtain short sequences from one end (single-end sequencing) or both ends (pair-end sequencing) while emitting photon signals during the sequencing process¹¹⁸. These photon signals were captured by a high-resolution camera. The reads are typically between 30 to 600 base pairs long, depending on the sequencing platform used¹¹⁸. Considerations for sequencing platform of choice depends on the application intended for the RNA-seq experiment¹²⁰. In quantitative gene expression analysis, it is more practical to sequence shorter reads at a higher depth with multiple biological replicates to improve the statistical power of the RNA-seq data¹²⁰. In RNA-seq experiments intended to be used for de-novo transcriptome assembly or for detecting splice variants between samples, longer read lengths are critical to achieving a high quality assembly. Following sequencing, the resulting reads are either aligned to a reference genome or assembled de novo (in species without reference genome) to produce a genome-scale transcription map that consists of both the transcriptional structure and/or level of expression for each gene^{120,126}.

Many bioinformatics tools exist for analysing RNA-seq data. In RNA-seq analysis, multiple software packages are required, as each software is catered for

different intended purposes^{127,128}. The first step for NGS-based methods for analysis is an alignment step, where the short reads from NGS are aligned against a reference genome to produce an alignment file¹²⁸. This file can be in either SAM format or the binary BAM format. The major differences between RNA-seq analysis and DNA sequencing are the need for splice adjustment¹²⁸. As the DNA resequencing are generated from the whole genome while RNA-seq is generated from mature mRNA or ribosomal RNA depleted total RNA, splicing events need to be accounted for. Some of the highly cited RNA-seq reads mapping softwares that are splicing sensitive include TopHat2¹²⁹ and STAR¹³⁰.

However, in instances when no reference genome or reference transcriptome is available, transcriptome de-novo assemblies are necessary¹²⁶. To perform transcriptome de-novo assemblies, one popular software suite used is Trinity¹³¹. Trinity partitions sequencing reads into individual de-Brujin graphs, with each de-Brujin graph representing the transcriptional complexity at a given gene or locus¹²⁶. Trinity then processes each graph independently to extract full-length splicing isoforms and to tease apart transcripts derived from paralogous genes. The outcome assemblies can be used as a reference to map the sequencing reads for downstream analysis. The second step for RNA-seq analysis is the differential gene expression analysis^{125,127,128}. This step is to identify the genes that are differentially expressed between two or more groups of samples. There are two types of methods: count based and gene length adjusted methods¹²⁷. Sequencing readcount based softwares includes DESeq¹³² and edgeR¹³³ while length adjusted methods include Cufflinks¹³⁴. These softwares allow the quantification of gene expression and comparison between the groups to identify the differentially expressed genes/transcripts. The third step of RNA-seq analysis

574 involves classifying the genes, that are differentially expressed in the two groups
575 examined, into their respective gene ontologies and pathways involved¹³⁵. By
576 classifying them into groups of gene functions, a more informative overview of
577 gene functions for the differentially expressed genes can be obtained.

578 **2.4.3 Estimation of evolution rates of coding genes**

579 Genes exhibiting sex-biased expression profiles exhibit a higher rate of evolution
580 compared to unbiased genes, possibly due to sexual selection and/or natural
581 selection⁹⁹. One way to investigate the evolution rate of a gene is by investigating
582 the synonymous to non-synonymous substitution ratio (dN/dS) of a gene's coding
583 region¹³⁶. By comparing the dN/dS of gene orthologues between closely related
584 species, one may estimate if a gene is under positive selection, neutral selection
585 or purifying selection^{137,138}.

586 Non-synonymous to synonymous mutation rate (dN/dS), also known as
587 Ka/Ks ratio, is used to quantitatively estimate the balances between beneficial
588 mutations, neutral mutations or purifying selections.¹³⁹ This simple calculation
589 takes advantage of the degeneracy of codons for protein translation. In a triplet
590 codon system, redundancies of the genetic codes allow for mutations to
591 accumulate on the exons without causing changes in protein translation¹³⁷. Such
592 mutations are known as synonymous substitution or mutations. On the other hand,
593 some genetic mutations on exons may result in protein sequence changes, known
594 as non-synonymous substitution or mutation. In a scenario when a beneficial
595 mutation resulting in amino acid changes (non-synonymous substitution) confers
596 genetic fitness to a gene, this gene is under positive selection^{138,140}. On the other

597 hand, when a gene can only accumulate mutations on the nucleic acid scales
598 without causing amino acid changes (synonymous substitution), the mutation
599 does not translate into a new amino acid sequence. Non-synonymous mutations
600 are rare and are usually lethal as it creates a non-functional protein structure.
601 Synonymous mutations are more common and are less subjective to selection
602 pressure¹³⁸.

603 Estimation of the evolution rate between species may be performed by
604 comparing the ratio between non-synonymous mutation rate and synonymous
605 mutation rate^{137,141}. When dN/dS is higher than 1, I may infer that the amino acid
606 changes are under favourable selection pressure, allowing the accumulations of
607 more non-synonymous mutations¹⁴¹. When a dN/dS is close to zero, it implies
608 that the gene function is highly susceptible to amino acid changes and non-
609 synonymous mutations are under purifying selection pressure¹⁴¹. Several
610 software suites are available for the calculation of dN/dS ratio. The most
611 commonly used software suite is the Phylogenetic Analysis by Maximum
612 Likelihood (PAML) package¹³⁶. Others such as the DataMonkey suite, take into
613 account site-specific dN/dS ratio rather than the dN/dS ratio of the entire coding
614 sequence, as the latter may mask important positive selection events in a very
615 long gene¹⁴².

616

617 **2.5 Overview of genome editing approaches to functionally** 618 **characterize genes in teleosts**

619 One important analysis to understand a gene's function is to perform a loss-of-
620 function analysis^{143,144}. This can be achieved by several means including N-ethyl-

621 N-nitrosourea (ENU) mutagenesis, zinc finger nucleases, transcription activator-
622 like effector nucleases (TALENs) and Clustered Regularly Interspaced Short
623 Palindromic Repeats (CRISPR/Cas9) techniques^{143,144}. Many mutant lines from
624 popular teleost models such as *Danio rerio* and *Oryzias latipes*, were produced
625 by random SNP-inducing mutagenesis using ENU¹⁴⁴. Many of these SNPs create
626 a premature stop codon in the coding sequence of a gene, resulting in a truncated
627 non-functional gene. However, this technique produces random mutations along
628 the genome, hence it is not a technique of choice when the mutation needs to be
629 on a specific location¹⁴⁴.

630 Zinc finger nucleases are one of the first true genome editing techniques
631 available for researchers to create site specific mutation on a genome¹⁴³. They are
632 made up of about thirty amino acids that can bind to nucleotide triplets, and all
633 64 possible trinucleotide combinations have been designed^{143,145}. These zinc
634 fingers can be chained together through molecular cloning and conjugated to a
635 FokI nuclease to cleave target DNA sequences with a set of nucleotide triplets
636 nearby¹⁴⁵. Another technique similar to zinc finger nuclease is TALEN, in which
637 proteins with DNA binding motifs directs FokI nuclease to cleave DNA at a
638 specific site in the genome¹⁴⁶. The advantage of TALEN over zinc finger
639 nucleases is that it recognizes single nucleotides instead of DNA triplets¹⁴³. This
640 allows for an easier designing process, without the limitation of a DNA triplet
641 requirement. TALEN also showed higher accuracies in comparison to
642 CRISPR/Cas9 with lower off-targets¹⁴³.

643 The latest genome-editing technique available is the CRISPR/Cas9
644 system¹⁴⁷. This technique takes a very different approach compared to zinc finger

nuclease and TALENs^{143,144}. It relies on two components, one being the Cas9 protein which cleaves the target DNA and the other being the site-specific guide RNA¹⁴⁷. The guide RNA is designed to be site-specific and upstream of a PAM-site containing the DNA motif “-NGG”¹⁴⁷. This technique is now employed in various mammalian cell lines and in-vivo in various organisms including non-model organisms¹⁴⁷⁻¹⁴⁹. The ease of synthesizing single stranded guide RNA in-vitro allows for multiplex CRISPR/Cas9 to cleave multiple sites on a targeted gene. This allows for a higher probability of a gene knockout event.

2.5.1 *In vivo* CRISPR/Cas9 mutagenesis in teleosts

Due to the simpler workflow and lower cost as compared to ZFN and TALENs, laboratories working on model fish species such as zebrafish and medaka began adopting CRISPR/Cas9 techniques to generate site-specific knockout mutations for gene functional analyses¹⁵⁰⁻¹⁵². Besides zebrafish and medaka, CRISPR/Cas9 protocols have been developed for various non-model species including *A. burtoni*¹⁵³, Atlantic salmon (*Salmo salar*)¹⁵⁴, Nile tilapia (*O. niloticus*)¹⁵⁵, Red sea bream (*Pagrus major*)¹⁵⁶ and channel catfish (*Ictalurus punctatus*)¹⁵⁷. To perform genome editing *in vivo*, a mixture of Cas9 protein or mRNA with the intended gRNA(s) are co-injected into single-cell stage embryos using a micromanipulator attached to a glass needle pipette^{149,152}. Various online gRNA designing softwares are available for estimating the off-target effect and the efficiency of a gRNA^{158,159}. Site-specific gRNA templates can be cloned into a vector and linearized as a transcriptional template. Alternatively, a cloning-free method has been reported to use PCR amplifications to generate gRNA templates, which allows for an even shorter time for gRNA synthesis¹⁵⁰. Cas9 can be

669 delivered as mRNA or by using commercially available Cas9 proteins conjugated
670 with nuclear localisation signal.

671 The use of an *in vivo* model for CRISPR/Cas9 experiments allows for the
672 phenotypic observation of body development when a target gene has been
673 knocked out^{149,150}. Should a gene be indispensable for development, a conditional
674 knockout may be performed. However, *in vivo* Cas9 experiments tend to generate
675 mosaic F0 animals as the cut-sites are repaired in a random fashion using non-
676 homologous end joining (NHEJ) repair mechanisms^{149,150}. Subsequent crossings
677 are necessary to generate stable homozygous knockout lines^{149,150}.

678 **2.5.2 *In vitro* CRISPR/Cas9 in teleost cell lines**

679 *In vitro* CRISPR/Cas9 genome editing has been reported for some fish cell lines
680 including Chinook salmon (*Oncorhynchus tshawytscha*) CHSE cells¹⁶⁰ and
681 medaka cell lines¹⁶¹. Due to the majority of CRISPR/Cas9 resources available on
682 plasmid repositories such as Addgene being tailored for mammalian cell culture
683 genome editing, *in vitro* genome editing in fish cell lines has been met with
684 several difficulties. Firstly, all-in-one plasmids used in mammalian cells genome
685 editing does not work efficiently on fish cells as the polymerase III promoter used
686 in these plasmids is non-functional in fish cells¹⁶¹. Secondly, the transfection
687 efficiency to introduce sgRNAs along with Cas9 mRNAs has been low, likely
688 due to the difficulty in transfecting fish cells and the low stability of RNAs in
689 fish cell culture medium. Nevertheless, successful genome editing has been
690 reported with the use of ribonucleoprotein delivery, which is co-transfection of
691 Cas9 protein with the intended sgRNAs¹⁶¹. The mutation efficiency has been

692 reported to be up to 61.5% with this method in medaka fish cells¹⁶¹. This method
693 may allow for the investigation of loss-of-function mutations when *in vivo*
694 genome editing is not feasible (e.g. gene loss-of-function is lethal during
695 embryogenesis).

696 **2.6 Knowledge gap in understanding the molecular mechanisms of sexual** 697 **size dimorphism in tilapia**

698 In this section, I would like to address some of the knowledge gaps in
699 understanding the epigenetic and genetic architecture of sexual size dimorphism
700 in tilapia.

701 Firstly, studies of sex-biased genes in somatic tissues are lacking in tilapia.
702 Phenotypic plasticity is well correlated with gene expression levels in an
703 organism. Understanding which genes are sex-biased in expression may help us
704 narrow down the candidate genes that may be responsible for the traits of sexual
705 size dimorphism in tilapia.

706 Secondly, the CpG methylation landscape in the tilapia genome is unknown.
707 No single-base resolution epigenome has been reported in tilapia. Epigenetic
708 changes have been reported to be responsible for gene expression differences
709 between individuals. Differentially methylated regions may exist within the
710 tilapia genome. Identification of these differentially methylated regions may
711 provide candidate cis-regulatory regions for studying sexual size dimorphism in
712 tilapia.

713 Thirdly, sex-biased genes have been shown to be under an accelerated rate
714 of evolution in other species, however, this hypothesis has not been tested in
715 tilapia before. Sexual size dimorphism may have been a consequence of sexual

716 selection or/and natural selection. Estimation of the evolution rate of sex-biased
717 genes may provide clues on candidate genes that have been under selection
718 pressure for the trait of sexual size dimorphism.

719 Besides the abovementioned, the causative factor(s) leading to sexual size
720 dimorphism remains unknown. Sexual size dimorphism may have been promoted
721 by sex hormones or sex-biased genes. It is also possible that it is a cumulative
722 effect of both sex hormones and sex-biased genes acting together in synchrony
723 to promote growth in males. These two factors may also divert energy
724 homeostasis in female tilapia towards the production of tilapia ovum, as female
725 gametes in tilapia are larger and energetically costlier to produce than male
726 gamete⁸².

727 **2.7 Aims and objectives of this thesis**

728 The purposes of this thesis are two-fold: (1) to explore the genetic and epigenetic
729 architecture in male and female tilapia to understand more about the mechanisms
730 underlying the sexual size dimorphism in tilapia, and (2) to identify DNA
731 markers associated with higher growth rate to facilitate the selection of fast-
732 growing tilapia. Although several studies have mapped the QTL for growth and
733 sex in tilapia, the transcriptomic profile differences of male and female tilapia in
734 somatic tissues were not yet examined. Also, the DNA methylation landscape
735 differences between male and female tilapia are not known. This thesis aims to
736 identify the sex-biased gene expressions between male and female tilapia in
737 somatic tissues and to explore the differences in epigenetic landscapes between
738 the two sexes.

739 I hypothesized that masculinization or feminization of somatic tissues such
740 as the brain and skeletal muscles occur during the early stages of sex
741 differentiation in tilapia. Brain sexual differentiation during early stages of
742 development may drive differential expression of neuroendocrine peptides which
743 may lead to differential growth rates between males and females^{72,74,162-165}. These
744 neuroendocrine signals may include positive regulators of growth such as GH,
745 GHRH, Agouti-related protein (AgRP) and neuropeptide Y (NPY)^{83,166-172}.
746 Negative regulator of growth may also cause sexual size dimorphism by slowing
747 down the growth rate in female tilapia. Potential negative regulators of growth
748 includes somatostatin, which is an inhibitor of GH secretion^{173,174}. This thesis
749 aims to identify sex-biased genes through RNA-seq experiment and characterize
750 the function of candidate sex-biased gene in relation to sexual size dimorphism
751 in tilapia.

752 Besides the above hypothesis, the effects of masculinization or feminization
753 may include epigenetic changes and/or sex-biased gene expressions between the
754 sexes^{61,63,104}. As growth is a quantitative complex trait, it is hypothesized that the
755 numbers of sex-biased genes will be large, with each gene contributing small
756 phenotypic differences in growth between male and female tilapia. Besides gene
757 expression differences, I hypothesized that epigenetic regulation, in form of CpG
758 methylation differences can regulate the expression levels of sex-biased genes in
759 tilapia. Thirdly, sexual size dimorphism is a form of phenotypic plasticity. A
760 relaxed purifying selection has been suggested as a pre-requisite for phenotypic
761 plasticity to evolve^{112,115,116}. Hence, I hypothesized that the sex-biased genes

762 identified from our RNA-seq data are under relaxed purifying selection and/or
763 positive selections in comparison with non-biased genes.

764 Besides identifying possible causative genes for sexual size dimorphism in
765 tilapia, this thesis also aims to identify SNPs that can be applied for marker
766 assisted selection in tilapia breeding¹⁷⁵. Sex-biased genes that are related to
767 growth will be genotyped for polymorphic DNA markers such as SNPs, indels
768 or microsatellites. Association studies between the identified DNA markers and
769 phenotypes including bodyweight and sex will be conducted to select candidate
770 SNPs for future marker assisted selection program.

771 Detailed objectives of this thesis are listed as follows:

- 772 • Generate a transcriptomic dataset of sex-biased gene expression in tilapia
773 somatic tissue.
- 774 • Produce a single base resolution whole genome bisulfite sequencing
775 dataset in a tilapia somatic tissue.
- 776 • Identify the sex-biased genes expressed in somatic tissues of male and
777 female tilapia.
- 778 • Detect regions of sexually dimorphic DNA methylation in the genome of
779 tilapia.
- 780 • Examine the effect of CpG methylations on sex-biased gene expressions.
- 781 • Investigate the rate of evolution of genes with sex-biased expression in
782 somatic tissues.

- 783 • Analyze the function of a candidate gene using CRISPR/Cas9 genome
784 editing.
- 785 • Discover DNA markers associated with increased growth rate or sex
786 phenotype for marker-assisted selection of fast-growing tilapia.
- 787 To achieve the above-mentioned objectives, the following methods and
788 procedures are utilized:
- 789 • Microsatellite sex genotyping of sexually immature tilapia.
- 790 • Collection of somatic tissues for RNA-seq and whole genome bisulfite
791 sequencing (WGBS) at different developmental time points.
- 792 • Construction and sequencing of ribosomal depleted total RNA libraries using
793 Illumina platforms.
- 794 • Sequencing bisulfite converted DNA samples from male and female tilapia
795 using Illumina platform.
- 796 • Mapping RNA-seq datasets using STAR-aligner.
- 797 • Mapping WGBS datasets using Bismark.
- 798 • Identifying sex-biased gene expression using DEseq2.
- 799 • Discivering pathways that are enriched in sex-biased genes.
- 800 • Detecting differentially methylated regions using methylKit.
- 801 • Inferring active cis-regulatory regions from WGBS datasets using
802 methylSeekR.

- 803 • CRISPR/Cas9 knockout of candidate genes in zebrafish and phenotypic
804 observations with relation to zebrafish growth rate.
- 805 • Identifying SNP markers associated with bodyweight and sex phenotype in
806 tilapia through association analysis.
- 807

810 Figure 3. Project roadmap outlined for this thesis.

811 **3 Materials and methods**

812 **3.1 Ethics statement**

813 All experiments in this study were approved by the IACUC Committee of the
814 Temasek Life Sciences Laboratory (Approval number: TLL(F)-11-001). The
815 experiments were performed according to the regulations and guidelines
816 established by this committee. Adult hybrid tilapia (*O. mossambicus* X *O.*
817 *niloticus*) used for skeletal muscle tissue collection were anaesthetized using
818 AQUI-S at 17 mg/L (AQUI-S, Lower Hutt, New Zealand) during visual sexing,
819 fin collection and traits measurement. Fish selected for skeletal muscle tissue
820 collection were euthanized with over dosage of AQUI-S (AQUI-S, Lower Hutt,
821 New Zealand) at 25 mg/L. For all fish used in brain tissue collection, aged 5 days
822 post hatch (dph), 30 dph and 90 dph, fish were euthanized in ice water slurry to
823 prevent possible effect of AQUI-S (AQUI-S, Lower Hutt, New Zealand) on brain
824 tissue transcriptome.

825

826 **3.2 Fish husbandry**

827 Mozambique tilapia used in this study were originally imported from South
828 Africa from wild populations and have been used for a saline tilapia aquaculture
829 breeding project project hosted by AVA, Singapore. A sub-population of the
830 Mozambique tilapia was maintained in the Marine Fish House facility in
831 Temasek Life Sciences laboratory. Fish were maintained in 200-litre fibre glass
832 tanks with circulatory fresh water at a constant flow rate of 2350 L/hour and with
833 constant aeration. Sexually matured fish were set up in mating pairs in 40-litre

834 glass tanks with circulatory water and observed for successful mating daily. For
835 the whole genome bisulfite sequencing and muscle transcriptome project, a male
836 *O. mossambicus* and a red Nile tilapia were used for crossing set up. For the brain
837 transcriptome project, both paternal and maternal brooders used were from *O.*
838 *mossambicus* line. Mating success can be observed when females are mouth
839 brooding. Tilapia embryos were carefully removed from the female's mouth at 2
840 days post fertilization. Embryos were kept in egg tumblers with slow constant
841 water flow at 3 L/hour.

842

843 **3.3 Traits measurement and visual sexing**

844 On the day before experiment, the tilapia was put on a fasting diet for one day.
845 Tilapia were anaesthetized with AQUI-S at a final concentration of 17 mg/L
846 before measurement. Body weight of each fish was weighed with digital scales
847 (Mettler Toledo, Ohio, United States) and recorded. Only sexually matured fish
848 were sexed through visual sexing. Tilapia with gonadal opening perpendicular to
849 the anterior-posterior body axis were sexed as females. Tilapia with gonadal
850 opening parallel to the body anterior-posterior axis were sexed as males. As a
851 reference, a cartoon describing the phenotype of tilapia genital openings can be
852 found at Figure 4.

853

854

Figure 4. Cartoon showing the phenotypic differences between male and female tilapia genital organs¹⁷⁶.

3.4 Tissue collection

Fin clips were collected from each tilapia and stored in absolute ethanol at -20°C until DNA isolation. For muscle whole genome bisulfite sequencing, skeletal dorsal muscle was dissected, minced and divided into two portions, one for DNA isolation and another for RNA isolation. Samples for DNA isolation were stored in absolute ethanol while samples for RNA isolation were snap frozen in liquid nitrogen and kept at -80°C until RNA extraction.

For brain transcriptome analysis samples, 5 days post hatch (dph), 30 dph and 90 dph Mozambique tilapia were used. Fish were first anaesthetized in ice slurry. Each immobile fish was decapitated and an incision was made on the skull area without damaging the brain tissue. Using a stainless-steel fine tweezer, brain tissue was dissected out of the skull under a stereomicroscope (Leica, Wetzlar,

871 Germany) and stored in 300 µl of RNAlater (Thermo Fisher Scientific,
872 Massachusetts, United States) at -80°C until ready for RNA isolation. The
873 remaining trunk body of each fish was stored in absolute ethanol at -20°C until
874 ready for DNA isolation for sex genotyping.

875 **3.5 Molecular sexing with microsatellite markers**

876 DNA isolation for sex genotyping was carried out using a method described by
877 Yue and Orban¹⁷⁷. Four microsatellites were used, namely S1542, S1543, S001
878 and S002, which are microsatellite markers flanking the sex determining region
879 in Mozambique tilapia as described by Liu et al⁴⁰. Sequences for these primers
880 are listed in Table 1. PCR for genotyping was performed using the following
881 thermal cycling conditions: 3 min at 95°C followed by 40 cycles of 30s at 95°C,
882 30s at 55°C and 1 min at 72°C and final step at 72°C for 10 min. To check PCR
883 results, 5 µl of PCR products was electrophoresed on 2% agarose gel stained with
884 0.001% ethidium bromide. Agarose gel containing DNA fragments was
885 visualized using Gel DOC XR (Bio-Rad, Hercules, USA). PCR products were
886 genotyped with capillary electrophoresis using a 3730xl DNA analyzer (Applied
887 Biosystems, California, United States). Alleles were called by comparing PCR
888 product fragment size with size standard GS-ROX-500 (Applied Biosystems,
889 California, USA) using the software GeneMapper (Applied Biosystems,
890 California, USA).

891

892

893 Table 1. Primers for genotyping microsatellite sex markers

Primer name	Primer sequence (5' – 3')
S1542F	TTGCATATTCCTCTATCCTCCTAA
S1542R	GATAGAAATACGCATGCAGAGTCC
S1543F	ACCGCGCTCTCAGCCTTCAT
S1543R	TGAGCGTCAGCGTCTCCATC
S0001F	TCTGGTGCATAGACATTCAACTG
S0001R	TTTTCTCCCCAAATTCATAAAGT
S0002F	TTTCTCCCCAGCTTGTGTAAGTCT
S0002R	TTGGACCTGGATTGCTGTTGATTA

894

895 3.6 Whole genome bisulfite sequencing

896 3.6.1 DNA isolation from skeletal muscle for whole genome 897 bisulfite sequencing

898 Skeletal muscles were cut into smaller pieces and lysed in 300 µl SET buffer
899 (0.4 M NaCl, 10 mM Tris-HCl pH 8.0, 2 mM EDTA pH 8.0 and 2% SDS) with
900 5 µg of proteinase K (Roche Life Sciences, Basel, Switzerland) in an orbital
901 shaker at 250 revolutions per minute (rpm) with temperature set to 55°C for 90
902 min or until complete digestion. Then, 0.5 mg of RNase A (Qiagen, Hilden,
903 Germany) was added into the lysate and incubated for 15 min at room
904 temperature. After that, 400 µl of 6 M NaCl was added into the sample mixture
905 and vortexed briefly until homogenous. The mixture was then centrifuged at 4°C
906 at 13,000 rpm for 30 min. The supernatant (approx. 600 µl) was transferred to a
907 new Eppendorf tube and 600 µl of isopropanol was added and mixed briefly. The
908 DNA-isopropanol mixture was precipitated overnight in a –20°C freezer and then
909 centrifuged at 4°C at 13,000 rpm for 30 min. The resulting DNA pellet was

910 washed twice with 80% ethanol and dissolved in Tris-EDTA pH 8.0 buffer and
911 stored at -80°C until ready for bisulfite conversion.

912 **3.6.2 Bisulfite conversion of genomic DNA**

913 Up to a total of 20 μl volume, 2 μg of each DNA sample was mixed with 1 ng of
914 unmethylated lambda DNA (Promega, Fitchburg, USA) as an internal control for
915 bisulfite conversion efficiency. The resulting mixtures were used for bisulfite
916 conversion on the ZYMO EZ-DNA Methylation-Gold kit (Zymo Research,
917 Irvine, USA). 130 μl of the CT Conversion Reagent was added to 20 μl of each
918 DNA sample in a PCR tube. Bisulfite conversion was performed on a PCR
919 machine with the following conditions: 98°C for 10 min, 64°C for 2.5 hours and
920 hold on 4°C . 600 μl DNA binding buffer (M-binding buffer), followed by DNA
921 conversion mixture, was added to spin column. Samples were centrifuged at 4000
922 x g for 5 min and the flow-through was discarded. 200 μl desulphonation buffer
923 (M-desulphonation buffer) was added into the column and allowed to stand at
924 room temperature for 20 min. Then the columns were centrifuged at 3000 x g for
925 5 min and the flow-through discarded. 400 μl of wash buffer (M-wash buffer)
926 was added and centrifuged at 3000 x g for 5 min with flow-through discarded.
927 Washing step was repeated again with centrifuge timing extended to 10 min. 15
928 μl of elution buffer was added into each column and incubated at room
929 temperature for 5 min. The columns were centrifuged at 3000 x g for 10 min and
930 the eluents were stored at -80°C .

931 3.6.3 Whole genome bisulfite sequencing (WGBS) library 932 construction

933 All converted DNA fragments were loaded onto individual lanes of a 1.5% Pippin
934 Gel Casette. Size selection was conducted on the Pippin-Prep platform (Sage
935 Science, Beverly, USA) with a target size in the range of between 300 to 320 bp.
936 DNA solutions were retrieved from the collection well of the gel cassette and
937 used for library construction using Illumina TruSeq DNA PCR-Free Library Prep
938 Kit (Illumina, San Diego, USA). End repair was conducted using 40 µl of end
939 repair mix (TruSeq ERP) and mixed by pipetting up and down slowly for 10
940 times. The PCR tubes were briefly centrifuged and incubated at 30°C for 30 min
941 on a PCR machine. 12.5 µl of A-tailing mix (TruSeq ATL) was added and mixed
942 by gentle pipetting. Samples were incubated on a PCR machine with heated lid
943 with the following program: 37°C for 30 min, 70°C for 5 min, 4°C for 5 min, and
944 then placed on ice. Sample of 2.5 µl was mixed by pipetting gently with 2.5 µl of
945 ligation mix (TruSeq LIG) and 2.5 µl of DNA adapters. The mixture was held at
946 30°C with a heated lid on a PCR machine for 10 min. Then, 5 µl of ligation stop
947 buffer (TruSeq STL) was added to each sample and mixed by pipetting. To purify
948 the library, 42.5 µl of sample purification beads (TruSeq SPB) were added and
949 mixed by pipetting. After 5 min of incubation at room temperature, the sample
950 mixture was placed on a magnetic stand for 5 min and all supernatant was
951 removed. Samples were washed twice with 80% ethanol while still on the
952 magnetic stand. Samples on the purification beads were air-dried for five min at
953 room temperature. 52.5 µl of resuspension buffer (TruSeq RSB) was then added
954 and mixed with pipetting. After 2 min of incubation at room temperature, samples

955 were placed on a magnetic stand for 5 min or until the solution was clear.
956 Supernatants were transferred to individual tubes and labelled as “CAP-sample-
957 X”. Then, in the second round of purification, 50 µl of sample purification beads
958 (TruSeq SPB) were added and mixed by pipetting. After 5 min of incubation at
959 room temperature, the sample mixture was placed on a magnetic stand for 5 min
960 and all supernatant was removed. Samples were washed twice with 80% ethanol
961 while still on the magnetic stand. Samples on the purification beads were air-
962 dried for five min at room temperature. 22.5 µl of resuspension buffer (TruSeq
963 RSB) was then added and mixed with pipetting. After 2 min of incubation at
964 room temperature, samples were placed on a magnetic stand for 5 min or until
965 the solution was clear. 20 µl of eluents were transferred to a clean tube and
966 labelled as final target library.

967 To quantitate the library concentration of each library using KAPA
968 Library Quantification Kits for Illumina Platforms, samples were diluted with
969 DNA dilution buffer (10 mM Tris-HCl, pH 8.0, 0.05% Tween® 20) at a dilution
970 rate of 1:10,000 and 1:100,000. For the qPCR reaction mix setup, 10 µl of KAPA
971 SYBR FAST qPCR Master Mix (2 x), 2 µl of Primer Premix, 4 µl of PCR-grade
972 water and 4 µl of the diluted library were mixed. Three technical replicates were
973 setup for each library concentration and also three each for six DNA standards
974 provided by the KAPA Library Quantification Kits. qPCR reaction was
975 conducted in BioRAD CFX96 using the following thermocycling program:
976 initial denaturation at 95°C for 5 min, 35 cycles of denaturation at 95°C for 30s
977 followed by 60°C annealing, extension and data acquisition for 45 seconds and
978 finally a melt curve analysis from 65°C to 95°C.

979 Library concentration was then calculated using a template Excel worksheet
980 provided by KAPA ([http://www.kapabiosystems.com/document/kapa-library-](http://www.kapabiosystems.com/document/kapa-library-quantification-data-analysis-template/?dl=1)
981 *quantification-data-analysis-template/?dl=1*). The final library was then
982 sequenced on Illumina HiSeq 2000 (Illumina, San Diego, USA) with 2 x 100 bp
983 paired-end protocol by BGI-Shenzhen (BGI-Shenzhen, Shenzhen, China).

984 **3.6.4 WGBS bioinformatic analysis**

985 **3.6.4.1 Reads quality control and mapping of bisulfite sequencing reads**

986 FASTQ reads were subjected to quality control using the IlluQC.pl module from
987 NGSToolKit Version 2.3. The clean reads were aligned to the reference genome
988 oreNil214 using Bismark v0.14.423. The Nile tilapia reference genome was
989 downloaded from GenBank at the accession number GCA_001858045.3. The
990 two strands of the reference genome (Watson strand and Crick strand) were
991 modified in silico to convert all C's to T's, using the
992 **bismark_genome_preparation** command with indexing format following
993 Bowtie2 requirement. Reads were mapped to these two modified genomes using
994 Bismark with the following parameters: **-q -p -N 1 --bowtie2**. The resulting
995 BAM files were sorted and converted to SAM format using Samtools V1.444.
996 This process was repeated using GenBank assembly J02459.1 for lambda phage
997 reference genome to map spiked-in unmethylated cl857 Sam7 Lambda DNA for
998 quantifying bisulfite conversion efficiency.

999 **3.6.4.2 Differentially methylated region analysis**

1000 Differentially methylated regions (DMRs) analysis was conducted in the R-
1001 environment using the R-package methylKit. BAM files produced from the

1002 previous step were read into R via the **processBismarkAln** command in
1003 methylKit. Each sample was then filtered out of low coverage bases and extreme
1004 coverage base position with the command **filterByCoverage** with minimal
1005 coverage of 10X and removing the top 0.1% base positions with the highest
1006 coverage. Single base resolution differential analysis was performed using the
1007 command **calculateDiffMeth** in methylKit V3.8 with the following parameters:
1008 slim=TRUE, weighted.mean=TRUE, num.cores=12. Tiling windows to identify
1009 DMRs of 1000bp windows were defined with **tileMethylCounts** command with
1010 the parameters win.size=1000 step.size=1000. Subsequently, DMR analysis was
1011 conducted using methylKit in the R-environment. The parameters for
1012 calculateDiffMeth in methylKit were as follows: slim=TRUE,
1013 weighted.mean=TRUE, num.cores=12.

1014 RepeatMasker was used to annotate the repeat elements found in the
1015 tilapia reference genome¹⁷⁸. The parameters used for RepeatMasker were as
1016 follows: -pa 5 -s -gff -species vertebrates. The generated GFF files were used to
1017 annotate the tilapia genome for calculation of methylation profiles across various
1018 repeat elements. Calculation of DMRs in repeat elements was conducted with
1019 methylKit using the same protocols as described above but with windows defined
1020 by annotations of repeat elements annotated by RepeatMasker.

1021

1022 **3.6.4.3 Active cis-regulatory region analysis**

1023 MethylSeekR was used to infer a regulatory element landscape in the R-
1024 environment¹⁰³. Biostrings-based reference genome data were required for
1025 MethylSeekR but were unavailable for the Nile tilapia reference genome as it is

not a model species. Hence, a Biostrings-based R library, BSgenome.Oniloticus.UCSC.oreNil2.tar.gz, was created for oreNil2 using BSgenome in the R-environment. Subsequently, unmethylated regions (UMRs) and low-methylated regions (LMRs) were identified using MethylSeekR with the Partially Methylated Domains (PMDs) masked. An arbitrary promoter region was defined as 500 bp upstream of the transcriptional start site for each transcript. UMRs and LMRs were mapped to their respective coordinates on SeqMonk (<http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>). All arbitrary promoter regions were scored and quantitated based on an enrichment test for LMRs on SeqMonk. Motifs enrichment analyses were conducted on the MEME suite¹⁷⁹.

3.7 Transcriptome analysis

3.7.1 RNA isolation from skeletal muscle and brain tissue for transcriptome analysis

RNA used for all transcriptome analysis in this thesis was extracted using TRIzol (Thermo Fisher Scientific, Massachusetts, United States) following the standard manufacturer's protocol. Samples were ground in 300 µl of TRIzol with a handheld homogenizer (Argos Technologies, Illinois, United States). Subsequently, 700 µl of TRIzol was added and incubated at room temperature for 5 min. 200 µl of chloroform was added into the samples and vortexed rigorously for 30 seconds and then incubated at room temperature for 3 min. Samples were centrifuged at 15,000 rpm at 4°C. Five hundred µl of the aqueous

1049 layer was mixed with 500 µl isopropanol with 1 µl of glycogen (Roche, Basel,
1050 Switzerland). Samples were vortexed briefly and precipitated at -20°C for
1051 overnight. Samples were centrifuged at 15,000 rpm for 15 min at 4°C. The
1052 supernatant was discarded and 1000 µl of 70% ethanol was added to wash the
1053 RNA pellet. Samples were centrifuged at 15,000 rpm for 15 min at 4°C. The
1054 supernatant was discarded by careful pipetting and the resulting pellet was air-
1055 dried at room temperature. Thirty µl of RNase-free water was then added into
1056 each tube and incubated at 55°C for 10 min or until the pellet was no longer
1057 visible. Samples were stored in a -80°C freezer until ready for library
1058 construction.

1059 **3.7.2 mRNA sequencing library construction**

1060 One µg of total RNA was diluted in RNase-free water to a final volume of 50 µl.
1061 Fifty µl of oligo-dT RNA purification beads were added into each tube and mixed
1062 gently by pipetting. Sample tubes were incubated at 65°C on a PCR machine for
1063 5 min to denature the mRNA. Samples tubes were incubated at room temperature
1064 for 5 min to allow the binding of poly-A tail mRNA onto the magnetic beads.
1065 Tubes were placed on magnetic stands for 5 min at room temperature and the
1066 supernatant was discarded from each tube. Tubes were removed from magnetic
1067 stands and 200 µl of beads washing buffer (TruSeq RNA BWB) was added and
1068 mixed by gentle pipetting. The tubes were returned to the magnetic stand for 5
1069 min and the supernatant was removed. The tubes were removed from the
1070 magnetic stand and 50 µl of elution buffer was added with gentle mixing by
1071 pipetting. The tubes were incubated at 80°C for 2 min on a PCR machine and
1072 returned to room temperature. Fifty µl of bead binding buffer (TruSeq RNA BBB)

1073 was added into each tube and incubated at room temperature for 5 min. Tubes
1074 were placed back onto the magnetic stand for 5 min and all supernatant was
1075 removed. Beads were washed with 200 μ l of beads washing buffer and returned
1076 to the magnetic stand for 5 min. All supernatant was removed and 19.5 μ l of
1077 elution buffer was added into each tube, mixed by pipetting and incubated at
1078 room temperature for 5 min. Samples were placed onto a PCR machine at 94°C
1079 for 8 min and returned to ice.

1080 Sample tubes containing the magnetic beads were placed on magnetic
1081 stands for 5 min. Seventeen μ l of the supernatant (fragmented and primed mRNA)
1082 were mixed by pipetting with 8 μ l of First Strand Master Mix (TruSeq RNA First
1083 Strand Master Mix) containing SuperScript II enzyme. Samples were incubated
1084 in a PCR machine with the following protocol: 55°C for 10 min, 42°C for 50 min
1085 and 70°C for 15 min and hold at 4°C. 25 μ l of second strand master mix (TruSeq
1086 RNA SSM) was added into each tube, mixed gently by pipetting, and placed on
1087 PCR machine to incubate at 16°C for 1 hour. Then, 90 μ l of AMPure XP beads
1088 was added into each tube containing 50 μ l of double stranded cDNA and mixed
1089 gently by pipetting. Tubes were incubated at room temperature for 15 min and
1090 placed onto magnetic stands for 5 min at room temperature. Up to 135 μ l of
1091 supernatant was removed and discarded from each tube and 200 μ l of 80%
1092 ethanol was added into each tube to wash the beads. After 30 seconds of
1093 incubation, 80% ethanol was removed and washing steps were repeated again.
1094 Sample tubes were air-dried while still on the magnetic stands for 15 min at room
1095 temperature. Tubes were removed from the magnetic stand and 52.5 μ l of
1096 resuspension buffer (TruSeq RNA RSB) was added into each tube and mixed by

1097 pipetting. After incubation at room temperature for 2 min, tubes were returned to
1098 the magnetic stands and incubated at room temperature for 5 min. Fifty µl of the
1099 supernatant was transferred to a new tube for end repair processes. Ten µl of
1100 resuspension buffer (TruSeq RNA RSB) and 40 µl of end repair mix (TruSeq
1101 RNA ERM) were added into each tube and mixed by pipetting. Tubes were
1102 incubated at 30°C on a PCR machine for 30 min. 160 µl of AMPure XP beads
1103 were added into each tube and mixed by pipetting. Tubes containing double
1104 stranded cDNA and AMPure XP beads were placed on magnetic racks for 15 min
1105 at room temperature or until the liquid was clear. All supernatant was removed
1106 from the tubes and 200 µl of 80% ethanol was added into each tube for washing.
1107 After incubation at room temperature for 15 min, the supernatant was removed
1108 and the washing step was repeated again. Beads in the tubes were allowed to air
1109 dry at room temperature for 15 min and removed from the magnetic rack.

1110 Resuspension buffer of 17.5 µl was added into each tube and mixed by
1111 pipetting. After incubation at room temperature for 5 min, tubes were returned to
1112 the magnetic stand until the liquid was clear. Clear supernatant (15 µl) was
1113 removed and placed in new tubes and 12.5 µl of A-tailing mix (TruSeq RNA
1114 ATL) was added into each tube and mixed by pipetting. Tubes were placed into
1115 PCR machine and incubated with the following program: pre-heat lid to 105°C,
1116 37°C for 35 min, 70°C for 5 min and hold at 4°C. Then, 2.5 µl of adaptor index
1117 (TruSeq RNA Adaptor Index) and 2.5 µl of ligation master mix (TruSeq RNA
1118 LIG) was added into the tubes containing A-tailed double stranded cDNA and
1119 mixed by pipetting. Tubes containing cDNA, adaptor index and ligation mix were
1120 incubated at 30°C for 15 min in a PCR machine and 5 µl of ligation stop buffer

1121 (TruSeq RNA STL) was added into each tube and mixed by pipetting gently. 42
1122 μ l AMPure XP beads were added into each tube and mixed gently by pipetting.
1123 Samples were incubated at room temperature for 15 min and placed on magnetic
1124 stands until the sample was clear. Up to 79.5 μ l of supernatant was discarded and
1125 200 μ l of 80% ethanol was added into each tube for 30 seconds and discarded.
1126 Wash steps were repeated once and beads were allowed to air dry on the magnetic
1127 stand at room temperature for 20 min. Tubes were removed from the magnetic
1128 stand and 52.5 μ l of resuspension buffer (TruSeq RNA RSB) was added into each
1129 tube and mixed with gentle pipetting. 50 μ l of AMPure XP beads were mixed
1130 into each tube by gentle pipetting and incubated at room temperature for 15 min.
1131 Tubes were placed on magnetic stands at room temperature until the liquid was
1132 clear. 95 μ l of supernatant was removed and discarded from each tube. 200 μ l of
1133 80% ethanol was added into each tube and discarded after 30 seconds. Wash step
1134 was repeated again and beads were allowed to air dry for 15 min at room
1135 temperature. Tubes were removed from the magnetic stand and 22.5 μ l of
1136 resuspension buffer (TruSeq RNA RSB) was added into each tube and mixed by
1137 pipetting. After 5 min of incubation at room temperature, tubes were placed on
1138 magnetic stands until the liquid was clear. Twenty μ l of supernatant was
1139 transferred into a new tube for PCR enrichment and 5 μ l of Illumina adaptor
1140 primer cocktail (TruSeq RNA PPC), 25 μ l of PCR master mix (TruSeq RNA
1141 PMM) were mixed in a new PCR tube. PCR cycling condition to enrich double-
1142 stranded cDNA were set as follows: initial denaturation at 98°C for 30 seconds,
1143 15 cycles of denaturation at 98°C for 10 seconds, annealing at 60°C for 30
1144 seconds and extension at 72°C for 30 seconds, followed by a final extension at
1145 72°C for 5 min and hold at 4°C.

1146 To clean up PCR products, 50 µl of AMPure XP beads were added into
1147 each PCR tube and mixed by gentle pipetting. PCR tubes with AMPure XP beads
1148 were incubated at room temperature for 15 min and placed on a magnetic rack
1149 until the solution was clear. Then, 95 µl of the supernatant was discarded and
1150 beads were washed twice with 200 µl of 80% ethanol. Beads were air dried at
1151 room temperature for 15 min. Resuspension buffer (TruSeq RNA RSB) of 32.5
1152 µl was mixed with the beads by pipetting and incubated at room temperature for
1153 5 min. Samples were placed on a magnetic rack until the liquid was clear and 30
1154 µl of the clear supernatant was transferred into a new tube and labelled as the
1155 final library.

1156 To quantitate the library concentration of each library using KAPA
1157 Library Quantification Kits for Illumina Platforms, samples were diluted with
1158 DNA dilution buffer (10 mM Tris-HCl, pH 8.0, 0.05% Tween® 20) at a dilution
1159 rate of 1:10,000 and 1:100,000. For the qPCR reaction mix setup, 10 µl of KAPA
1160 SYBR FAST qPCR Master Mix (2 x), 2 µl of Primer Premix, 4 µl of PCR-grade
1161 water and 4 µl of the diluted library were mixed. Three technical replicates were
1162 setup for each library concentration and also three each for six DNA standards
1163 provided by the KAPA Library Quantification Kits. qPCR was conducted in
1164 BioRAD CFX96 using the following thermocycling program: initial denaturation
1165 at 95°C for 5 min, 35 cycles of denaturation at 95°C for 30 seconds followed by
1166 60°C annealing, extension and data acquisition for 45 seconds and finally a melt
1167 curve analysis from 65°C to 95°C. Library concentration was then calculated
1168 using a template Excel worksheet provided by KAPA
1169 (<http://www.kapabiosystems.com/document/kapa-library-quantification-data->

1170 *analysis-template/?dl=1*). The final library was then sequenced on Illumina
1171 HiSeq 2000 (Illumina, San Diego, USA) with 2 x 100 bp paired-end protocol by
1172 BGI-Shenzhen (BGI-Shenzhen, Shenzhen, China).

1173 **3.7.3 Ribosomal depleted sequencing library construction**

1174 RNA samples were examined for RNA integrity check using Bioanalyzer RNA
1175 6000 Nano assay (Agilent, California, United States). Ribosomal depletion of
1176 RNA samples was conducted using the NEBNext® rRNA Depletion Kit
1177 (Human/Mouse/Rat) (New England Biolabs, Massachusetts, United States). One
1178 µg of total RNA was hybridized with DNA oligos which formed a
1179 complementary structure with ribosomal RNA. Hybridized samples were
1180 incubated with RNase H at 37°C for 30 min. DNase I and reaction buffer were
1181 added and incubated at 37°C for 30 min. RNA samples were purified with
1182 magnetic purification beads and stored in -80°C.

1183 Construction of stranded RNA-seq libraries was carried out using
1184 NEBNext Ultra Directional RNA Library Prep Kit for Illumina (E7420) (New
1185 England Biolabs, Massachusetts, United States). Five µl of ribosomal depleted
1186 RNA was used for each sample. RNA was fragmented for 15 min in cDNA
1187 synthesis buffer at 94°C. Random primers were added for first strand cDNA
1188 synthesis. After incubating at 42°C for 50 min, second strand synthesis master
1189 mix was added and incubated at 16°C for 1 hour. The cDNA cleanup was
1190 performed with 1.8 x volume of Agencourt AMPure XP beads. Library end repair
1191 and adaptor ligation were performed following the manufacturer's instruction.
1192 Twelve rounds of PCR cycles were performed to enrich the library and cleaned

1193 with Agencourt AMPure XP Beads (Beckman Coulter, California, United States).
1194 Libraries were quantified on BioRAD CFX96 using KAPA Library
1195 Quantification Kits (Roche, Basel, Switzerland). Libraries were mixed and
1196 sequenced on two separate runs using an Illumina NextSeq 500 (Illumina,
1197 California, United States).

1198 **3.7.4 Transcriptome bioinformatics analysis**

1199 **3.7.4.1 Reads quality control and mapping**

1200 Skeletal muscle transcriptome raw FASTQ reads were quality controlled using
1201 the IlluQC.pl module from NGSToolKit Version 2.3. The Nile tilapia reference
1202 genome was downloaded from GenBank at the accession number
1203 GCA_001858045.3¹⁸⁰ and used for the analysis of transcriptomes. The reference
1204 genome was then indexed for STAR alignment using the command **STAR --**
1205 **runMode genomeGenerate**¹³⁰. Good quality reads were mapped onto the Nile
1206 tilapia reference genome using the software STAR using default parameters¹³⁰.

1207 Brain transcriptome FASTQ data were cleaned with Trimmomatic V0.39
1208 with the following parameters: ILLUMINACLIP:TruSeq3-PE.fa:2:30:10
1209 LEADING:3 TRAILING:3 SLIDINGWINDOW:4:15 MINLEN:36¹⁸¹. Nile
1210 tilapia reference genome used was downloaded from GenBank at the accession
1211 number GCA_001858045.3¹⁸⁰. Reference genome was then indexed for STAR
1212 V2.7 alignment using the command **STAR --runMode genomeGenerate**¹³⁰. Good
1213 quality reads were mapped onto the Nile tilapia reference genome using the
1214 software STAR V2.7 and converted to sorted BAM files¹³⁰.

1215 **3.7.4.2 Differential gene expression analysis**

1216 To calculate the number of reads mapped to a feature (e.g. Gene), HTSeq-count
1217 were used¹⁸². The command to generate count table for each gene was: htseq-
1218 count -f bam -t gene -i Name --stranded=reverse sample.bam annotation.gtf >
1219 output.txt. Read count files were loaded into R using read.table function and
1220 samples were grouped accordingly to sex and developmental stages.
1221 Differentially expressed genes (DEGs) were calculated using DESeq2 in the R-
1222 environment to identify sex-biased genes between male and female tilapia
1223 samples. Sex-biased genes were defined as having adjusted *p* of less than 0.05
1224 and fold-change of more than two.

1225 **3.7.4.3 qPCR validation of transcriptome analysis in muscle tissue**

1226 To validate our muscle tissue transcriptome analysis, I selected six sex-biased
1227 candidate genes and designed primers using Primer3 with the following
1228 requirements: product size 220-300 bp and target annealing temperature 58°C.
1229 Four male and four female tilapia at 108 dph were euthanized in AQUI-S (AQUI-
1230 S, Lower Hutt, New Zealand) at overdose condition (175 mg/L). Skeletal muscle
1231 tissue was collected and snap frozen in liquid nitrogen until ready for RNA
1232 extraction. RNA extraction was carried out using TRIzol according to the
1233 manufacturer's protocol and eluted in RNase-free water. Reverse transcription
1234 was carried out using M-MLV reverse transcriptase with a mix of random
1235 hexamer primers and oligo-dT primers. cDNA products were diluted 10-fold with
1236 water prior to qPCR reaction. Real time PCR was conducted using KAPA SYBR
1237 Fast Universal Kit (Sigma-Aldrich, Missouri, United States) in CFX96 Touch™
1238 Real-Time PCR Detection System (Bio-Rad Laboratories, California, United

1239 States). qPCR data analysis was conducted using the delta delta Ct method.

1240 Primer sequences for qPCR analysis can be found in Table 2.

1241 Table 2. A list of primers used for qPCR validations in skeletal muscle tissue

Primer name	Primer sequence (5' – 3')
BMP1-q1F	GCAATGATGTGAGATTTGGGGT
BMP1-q1R	GGAGCACTCGTCCTTATCTGA
RAPGEF5-q1F	CAGCCAGTAATCAGTCGGCA
RAPGEF5-q1R	GTCACCTCTGGGGCTCACTTG
HOX11b-q1F	CAGAGCAAAAGCAGCAGTGG
HOX11b-q1R	AGGGGGTTCGTGGTGTAGTA
IRS2-q1F	ATGGCCAAGGGAGAGAGGAT
IRS2-q1R	ATCAACTCTCAAGCAGCCCC
MHC-q1F	GCTCACCTACCAGACTGAGG
MHC-q1R	TGTCTCTTACTCAGCACTTTCA
RGCC-q1F	GTTAGATAAAGACCTGAAACAGCG
RGCC-q1R	TGATGAACTTGGCTTGAGGT
Actin-q1F	TGACCCAGATCATGTTTCGAGAC
Actin-q1R	TGTGGTGGTGAAGGAGTAGCC

1242

1243 3.7.4.4 qPCR validation of brain sex-biased genes

1244 To validate our brain tissue transcriptome analysis, I selected 12 sex-biased
1245 candidate genes and designed primers using Primer 3 with the following
1246 requirements: product size 150-200, annealing temperature 61°C. Four male and
1247 four female tilapia at both 90 dph and 120 dph were euthanized in ice slurry with
1248 MS-222 (Sigma-Aldrich, Missouri, United States). Brain tissue was carefully
1249 dissected out of each fish and stored in RNAlater (Thermo Fisher Scientific,
1250 Massachusetts, United States) at -80°C. RNA extraction was carried out using
1251 TRIzol (Thermo Fisher Scientific, Massachusetts, United States) according to the
1252 manufacturer's protocol and eluted in RNase-free water. Reverse transcription
1253 was carried out using M-MLV reverse transcriptase (Promega, Fitchburg, USA)
1254 with a mix of random hexamer primers and oligo-dT primers. cDNA products

were diluted 10-fold with water prior to qPCR reaction. Real time PCR was conducted using KAPA SYBR Fast Universal Kit (Sigma-Aldrich, Missouri, United States) in CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, California, United States). qPCR data analysis was conducted using the delta delta Ct method. Primer sequences for qPCR analysis can be found in Table 3.

Table 3. List of primers for qPCR validations in brain tissue

Primer name	Primer sequence (5' – 3')
Actin-q1F	TGACCCAGATCATGTTTCGAGAC
Actin-q1R	TGTGGTGGTGAAGGAGTAGCC
pomc-q1F	CCTCTGCTGACGCTCTTCAA
pomc-q1R	GGGAACTCTTTGGTGCCTGA
GnRH3-q1F	GGCTTCCCCTGCAGTTCTAA
GnRH3-q1R	ATGGTTGCCTCAAGCTCTCC
AR-q1F	GCACACCAAGGTCAGCTTTC
AR-q1R	CAGAGTTGAAAATGTGGGGAAACA
LY75-qF	CAGGAGGCGAGGCTCTTTTG
LY75-qR	AGGGAATATGCGTCCAGGCT
IL21R -qF	CCGTCTGAGACATCAGTGGC
IL21R-qR	AGCTAAGCGGAAAACATCTCCA
ORC1-q1F	GGCAACGGAGATCTGTGAGC
ORC1-q1R	ATCCCAGCCGCCTAAACTCT
KDM7A-q1F	GACAGTAGGAAGATGGCGGC
KDM7A-q1R	CATCAGGGAGGGTCCGTGTA
4792-q1F	TACTGTCACGGTTGGAGGGT
4792-q1R	CTCCGTTGACGAAGGCAGTC
CXCR4-q1F	CTCTGCACAGCATGACAGGT
CXCR4-q1R	TCGAGGTGTCAGGTGCTGAT
ZMAT4-q1F	GTGCCTCTGCGGAAGAGTAAG
ZMAT4-q1R	TGCCACAATTGCAGAGGTGA
ncRNA1-qF	TGCCCAGGGGCATTACTGAT
ncRNA1-qR	GGCTTCTTCGTGGCTGCAAA
ncRNA2-qF	CATCACGATGCCGCCCC
ncRNA2-qR	TGCTAACATTTTGGTTGCTGTGA

1263 **3.7.4.5 Chromosomal locations of sex-biased genes**

1264 To obtain the chromosome location information of sex-biased genes, gene
1265 chromosome numbers were obtained by parsing gene annotation file (.gff file)
1266 into an R environment using the R package rtracklayer. The expected number of
1267 sex-biased genes per chromosome was calculated with the following formula:

1268 Expected # of genes on chr N =

$$1269 \frac{\text{\# of genes on chr } N}{\text{total number of genes}} \times \text{total number of sex biased genes}$$

1270

1271 **3.7.4.6 Gene ontology and pathway analysis**

1272 To obtain gene ontology pathway information, I BLAST searched Nile tilapia
1273 coding sequences to zebrafish coding sequences (downloaded from ENSEMBL)
1274 using blastn with the following parameters: **blastn -task dc-megablast -outfmt**
1275 **6 -max_target_seqs 1 -evaluate 0.001**. Zebrafish EntrezID and GO annotations
1276 were obtained from Bioconductor annotation package org.Dr.eG.db
1277 ([http://www.bioconductor.org/packages//data/annotation/html/org.Dr.eG.db.htm](http://www.bioconductor.org/packages//data/annotation/html/org.Dr.eG.db.html)
1278 l). Gene ontology enrichment analysis was performed using ClusterProfiler
1279 **enrichGO** function in R-environment¹⁸³. Zebrafish homologues of each sex-
1280 biased gene were searched in pantherDB zebrafish database for Panther Pathway
1281 enrichment¹⁸⁴.

1282

1283 **3.7.4.7 SNP calling from RNA-seq data**

1284 To extract single nucleotide polymorphisms (SNPs) from RNA-seq data, FASTQ
1285 files were mapped onto the reference genome using STAR-aligner with 2-pass

1286 protocol¹³⁰. Alignment was performed once to obtain the coordinates of all splice
1287 junctions. The output splice junction coordinates were used to re-index the
1288 reference genome with the following command: **STAR --runMode**
1289 **genomeGenerate --genomeDir \$genomeDir --genomeFastaFiles**
1290 **tilapia_referrence_genome.fa --sjdbFileChrStartEnd**
1291 **/path/to/splice/junction.out.tab --sjdbOverhang 75 --runThreadN 12**. The
1292 second round of alignment was conducted with STAR using the new genome
1293 index. Output BAM files were marked for duplicates and sorted into read groups
1294 using Picard (<http://broadinstitute.github.io/picard/>). Overhangs at exonic
1295 junctions were trimmed using GATK SplitNCigarReads and reassigned mapping
1296 quality¹⁸⁵. Base scores for reads were recalibrated with GATK
1297 BaseRecalibrator¹⁸⁵. SNPs and indels were called with samtools V1.9 mpileup
1298 function followed by bcftools V1.9 call function^{186,187}. After every library was
1299 called for SNPs and indels, all output variant call files (.vcf files) were merged
1300 using samtools merge function^{186,187}.

1301 **3.7.4.8 Non-synonymous and synonymous base substitution calculation**

1302 To estimate the evolution rate of sex-biased genes identified in this study, I
1303 downloaded coding sequences of closely related cichlid species from Ensembl,
1304 including *Maylandia zebra*, *Amphilophus citrinellus*, *Neolamprologus brichardi*,
1305 *Astatotilapia burtoni* and *Haplochromis nyererei*^{49,188}. All protein-coding sex-
1306 biased genes were used to calculate synonymous and non-synonymous mutation
1307 rates. Coding sequences were translated into amino acid FASTA files using
1308 **EMBOSS Transeq** and protein sequences of every species were concatenated
1309 into a single database.fasta file¹⁸⁹. Database.fasta file was then made into a

1310 BLAST database using the command: **makeblastdb -in database.fasta -dbtype**
 1311 **prot**¹⁹⁰. Tilapia protein sequences were BLAST against this database with the
 1312 following parameters: **blastp -query tilapia.fasta -out results.out -outfmt 6 -**
 1313 **evaluate 0.001 -seg yes -soft_masking true -max_target_seqs 5000**. Then, each
 1314 tilapia protein was grouped into orthologues groups using the python package
 1315 VESPA.py with the parameter **-e_value=0.001 -alignment_length=75 -**
 1316 **percent_identity=75**. Each homologous group of protein sequences were
 1317 subjected to multiple sequence alignment with MAFFT with the following
 1318 command: **mafft --auto --thread 12 input.fasta > alignment_output.mafft**¹⁹¹.
 1319 Protein alignment guided codon alignment was conducted using PAL2NAL with
 1320 the following command line: **pal2nal.pl alignment_output.mafft**
 1321 **coding_sequences.fasta -output paml -nogap > output.pal2nal**¹⁹². The output
 1322 file output.pal2nal was used as input alignment sequence for calculation of dN/dS
 1323 using codeML from the PAML software suite¹³⁶. Run parameters for codeML
 1324 were listed as:

1325 noisy = 0 * 0:quiet
 1326 verbose = 0 * 1:detailed output
 1327 runmode = -2 * -2:pairwise
 1328 seqtype = 1 * 1:codons
 1329 CodonFreq = 2 * 2:F3X4
 1330 model = 1
 1331 NSsites = 0
 1332 icode = 0 * 0:universal code

1333 $\text{fix_kappa} = 1$ * 1:kappa fixed, 0:kappa to be estimated
1334 $\text{kappa} = 1$ * initial or fixed kappa
1335 $\text{fix_omega} = 0$ * 1:omega fixed, 0:omega to be estimated
1336 $\text{omega} = 0.5$ * initial omega value

1337

1338 **3.8 Candidate gene functional analysis *in vitro***

1339 **3.8.1 OmB cell culturing**

1340 Cells used for in-vitro genome editing in this thesis were immortalized cell lines
1341 derived from the brain of a female Mozambique tilapia (OmB cell lines)¹⁹³. Cells
1342 were seeded at 10% confluency in 25 ml flasks containing 10 ml of Leibovitz's
1343 L-15 medium (Thermo Fisher Scientific, Massachusetts, United States)
1344 supplemented with 10% fetal bovine serum (FBS) and 1% Antibiotic-
1345 Antimycotic (100X) (Thermo Fisher Scientific, Massachusetts, United States).
1346 Flasks containing cells were sealed air tight with Parafilm and incubated at 28°C
1347 in an incubator. To passage cells, monolayer confluent OmB cells with more than
1348 70% confluency were washed three times with 1 x phosphate buffered saline
1349 (PBS) and 2 ml of trypsin solution was added. After 60 seconds of incubation at
1350 room temperature, flasks containing the mono cell layers were gently tapped
1351 twice and transferred into a new 10 ml Falcon tube with 5 ml of 1 x PBS to wash
1352 the cells. Cells were centrifuged at 150 g for 5 min to pellet the cells. The
1353 supernatant was removed in a biosafety cabinet and 5 ml of 1 x PBS were added
1354 to dislodge the cells. The cell wash step was repeated twice and fresh Leibovitz's

1355 L-15 medium containing 10% FBS and 1% antibiotic were added to dislodge the
1356 cell pellets and seeded at 10% confluency.

1357 **3.8.2 CRISPR/Cas9 plasmid for in-vitro genome editing of OmB cells**

1358 The plasmid used in the in-vitro genome editing of OmB cells were the
1359 pSpCas9(BB)-2A-GFP (PX458) plasmid from Addgene¹⁹⁴. This plasmid
1360 contains a cassette encoding Cas9 protein from *Streptococcus pyogenes* with 2A-
1361 EGFP and a cloning backbone for sgRNA. Guide RNA was designed using
1362 CRISPR RGEN Tool Cas-Designer. FASTA sequences of each exon were
1363 uploaded and designed for SpCas9 protein with the PAM motifs of 5'-NGG-3'.
1364 Oligos were ordered for each target specific region with a 5'-TAGG overhang
1365 for “sense” oligo and 5'-AAAC overhang for “antisense” oligo. Oligos were
1366 diluted to 100 µM with PCR-grade water prior to cloning. A phosphorylating and
1367 annealing mixture was assembled with the following composition: 1 µl sgRNA
1368 top (100 µM), 1 µl sgRNA bottom (100 µM), 1 µl T4 ligation buffer (10x), 1 µl
1369 T4 Polynucleotide Kinase and 6 µl PCR-grade water. Samples were mixed by
1370 gentle pipetting and incubated in a PCR machine with the following program:
1371 37°C for 30 min; 95°C for 5 min; ramp down to 25°C at 5°C min⁻¹.
1372 Phosphorylated double-stranded oligo was then diluted 1:200 with PCR-grade
1373 water prior to cloning. To clone double-stranded oligos into pSpCas9(BB)-2A-
1374 GFP plasmid, the following components were assembled : 100 ng of circular
1375 pSpCas9(BB)-2A-GFP circular plasmid, 2 µl diluted double-stranded oligo, 2 µl
1376 SmartCut buffer (10x), 1 µl DTT (10 mM), 1 µl BbsI endonuclease, 0.5 µl T7
1377 ligase and topped up to 20 µl. Samples were thermocycled on a PCR machine
1378 with the following protocol: 6 cycles of 37°C for 5 min followed by 21°C for 5

1379 min. Ligation products were transformed into competent *Escherichia coli* strains
1380 following standard heat shock procedure. Transformed cells were plated on LB
1381 plates containing ampicillin as a selective antibiotic and incubated overnight at
1382 37°C. Individual clones were picked and subcultured in 50 µl of LB containing
1383 ampicillin. 1 µl of subculture were used for colony PCR using the primers
1384 PX458-guideF and PX458-guideF which flanked the BbsI cut site in
1385 pSpCas9(BB)-2A-GFP plasmid. Sequencing of PCR product was conducted with
1386 BigDye sequencing chemistry on a 3730xl DNA analyzer (Applied Biosystems,
1387 California, United States). Sequencing results were analysed with Sequencher
1388 V5.1 (Gene Codes Corporation, Michigan, United States). Positive clones
1389 containing correct guide RNA template sequences were selected for culturing and
1390 for extraction of endotoxin-free plasmids using PureLink™ HiPure Plasmid
1391 Midiprep Kit (Thermo Fisher Scientific, Massachusetts, United States).

1392 **3.8.3 Transfection of OmB cells with Lipofectamine 3000**

1393 Prior to transfection, cells were seeded in 6 well plates with complete medium
1394 (Leibovitz's L-15 medium supplemented with 10% FBS and 1% antibiotics) and
1395 allowed to grow up to 70% confluency. Transfection of cells with pSpCas9(BB)-
1396 2A-GFP plasmids were performed with Lipofectamine 3000 (Thermo Fisher
1397 Scientific, Massachusetts, United States). Ten µl of Lipofectamine 3000 was
1398 mixed with 125 µl of Leibovitz's L-15 medium in one Eppendorf tube. 2.5 µg of
1399 plasmid DNA, 7.5 µl of P3000 reagent and 125 µl of Leibovitz's L-15 medium
1400 were mixed in another Eppendorf tube. Both Eppendorf tubes were mixed and
1401 allowed to incubate at room temperature for 15 min. After 15 min, DNA-lipid
1402 complex was added drop-wise into each well containing 1000 µl of Leibovitz's

1403 L-15 medium. The medium was replaced with complete medium (Leibovitz's L-
1404 15 medium supplemented with 10% FBS and 1% antibiotics) after incubation for
1405 24 hours at 28°C at ambient air. At 48 hours post transfection, cells were checked
1406 for GFP emission using an inverted microscope with a halogen lamp.

1407 **3.8.4 DNA extraction from transfected cells and T7 Endonuclease Assay**

1408 To validate genome editing at target cut sites, medium was removed from each
1409 well and cells were lysed with 300 µl SET buffer (0.4 M NaCl, 10 mM Tris-HCl
1410 pH 8.0, 2 mM EDTA pH 8.0 and 2% SDS) with 5 µg of proteinase K (Roche Life
1411 Sciences, Basel, Switzerland). Cells were dislodged and lysed by pipetting up
1412 and down each well for 20 times. Lysed samples were transferred into Eppendorf
1413 tubes and 400 µl of 5 M NaCl was added into the sample mixture and vortexed
1414 briefly until homogenous. The mixture was then centrifuged at 4°C at 13,000 rpm
1415 for 30 min. The supernatant (approx. 600 µl) was transferred to a new Eppendorf
1416 tube and 600 µl of isopropanol was added and mixed briefly. The DNA-
1417 isopropanol mixture was precipitated overnight in a -20°C freezer and then
1418 centrifuged at 4°C at 13,000 rpm for 30 min. The resulting DNA pellet was
1419 washed twice with 80% ethanol and dissolved in 100 µl of Tris-EDTA pH 8.0
1420 buffer.

1421 PCR primers flanking 500 bp around target sites were designed using
1422 Primer 3. PCR for target product flanking target sites were conducted with
1423 flanking primers with the following reagents: 1 U Taq polymerase, 0.3 µl dNTP
1424 (2.5 mM), forward flanking primer (10 mM), reverse flanking primer (10 mM),
1425 2.5 µl PCR buffer (10x), 100 ng of sample DNA and topped up with PCR-grade

1426 water to 25 µl. Samples were thermocycled in a PCR machine using the following
1427 protocol: 95°C for 3 min, 40 cycles of denaturing 95°C for 30 seconds, annealing
1428 55°C for 30 seconds and extension 72°C for 30 seconds, followed by 72°C for 5
1429 min and hold at 16°C. Prior to T7 endonuclease I digestion, samples were
1430 denatured at 95°C for 1 minute, followed by ramping at a rate of -0.1°C s⁻¹ to
1431 25°C. T7 endonuclease assay mixture composed of 1 U of T7 endonuclease I, 1
1432 µl of 10x NEB Buffer 2 and 8 µl of reannealed PCR product. Samples were
1433 incubated at 37°C for 1 hour and electropherised on a 2% agarose gel stained
1434 with 0.001% ethidium bromide.

1435 **3.8.5 Imaging of transfected cells**

1436 Before imaging of transfected cells, the complete medium was removed and
1437 washed once with 1 x PBS at room temperature. Two ml of 1 x PBS was added
1438 into each well for imaging. Cells were imaged on an inverted bright field
1439 microscope Leica DM IL LED (Leica, Wetzlar, Germany) with fluorescence
1440 filter for 395 nm emission. Image processing was conducted in ImageJ V1.52e.

1441 **3.8.6 Prediction of estrogen response elements in the promoter region of** 1442 **POMC**

1443 To understand what causes sex-biased gene expression of *POMC* in Mozambique
1444 tilapia, I focused on identifying and characterizing putative estrogen response
1445 elements (EREs) in the promoter or upstream region of *POMC* in tilapia. DNA
1446 sequence up to five kilo base pairs upstream of *POMC* transcription start sites
1447 were extracted from the *O. niloticus*_UMD_NMBU tilapia reference genome,
1448 uploaded on JASPAR website (<http://jaspar.genereg.net/>) and searched for

1449 ESR1, ESR2, ESRRRA and ESRRB binding motifs from the Vertebrata database.
1450 Search program used was the “Scan” function with a similarity matrix of at least
1451 85%.

1452 **3.8.7 Cloning of POMC ERE into Luciferase Vector**

1453 Two ERE corresponding to Estrogen Receptor Beta were amplified with PCR
1454 using the primers listed in Table 4. High fidelity PCR was conducted using the
1455 KAPA HiFi PCR Kit with the following PCR cycle program: 98°C for 3 min, 40
1456 cycles of 98°C for 30 seconds, 60°C for 30 seconds, 72°C for 90 seconds,
1457 followed by a final extension of 72°C for 10 min. PCR products were checked
1458 with gel electrophoresis on a 1% agarose gel stained with 0.001% ethidium
1459 bromide for successful amplifications. PCR products were cleaned with gel
1460 purification using FavorPrep Gel Purification Mini Kit (Favorgen Europe,
1461 Vienna, Austria). PCR fragments were double digested with restriction enzyme
1462 SacI and SmaI in 1 x CutSmart Buffer (New England Biolabs, Massachusetts,
1463 United States). Digested PCR fragments were cleaned again with gel purification
1464 using FavorPrep Gel Purification Mini Kit (Favorgen Europe, Vienna, Austria).
1465 To prepare a Luciferase expression vector for cloning, pGL3-Promoter Vector
1466 was digested with restriction enzyme SacI and SmaI in 1X CutSmart Buffer (New
1467 England Biolabs, Massachusetts, United States). Linearized vector fragments
1468 were cleaned again with gel purification using FavorPrep Gel Purification Mini
1469 Kit (Favorgen Europe, Vienna, Austria). Ligation of digested PCR fragment and
1470 linear pGL3-Promoter Vector were performed using T4 DNA Ligase (New
1471 England Biolabs, Massachusetts, United States). Ligated vectors were
1472 transformed into competent *Escherichia coli* cells using the heat shock method

for 45 seconds at 42°C using a water bath and plated on Ampicillin selection LB plates. Screening for positive clones was done using RVprimer3 and GLprimer2 through colony PCR. Clones containing PCR fragment size of approximately 1450 bp were selected for sub culturing in ampicillin LB media.

Table 4. Primers for cloning POMC estrogen response elements

Primer name	Primer sequence (5' – 3')
POMC_LucF	ATCGCGGAGCTCTTTTCCCACACAACCATCTCTCAGG
POMC_LucR	ATCGCGCCCGGGACCGTTAAATTGAGCTTAAACTTTGCA

3.8.8 Transfection of pomcERE-pGL3-promoter into OmB cells and Luciferase Assay

Endotoxin free plasmids used for cell transfection were extracted using PureLink™ HiPure Plasmid Midiprep Kit (Invitrogen, California, United States). Before OmB cells transfection with pomcERE-pGL3-promoter, OmB cells were plated on 24-well plates at 10% confluency on complete media. Transfection was performed when cells had grown to 70% to 90% confluency. Cell culture media were changed into L-15 media without FBS and antibiotic prior to transfection. For each well, 250 ng of Renilla control vector, 250 ng of pomcERE-pGL3-promoter vector, 1 µl of P3000 reagent, 1.5 µl of Lipofectamine 3000 and 50 µl of L-15 media were mixed and incubated at room temperature for 15 min. Transfection complex was added drop-wise into each cell culture well and incubated at 28°C for 24 hours. Culture media in control wells were replaced with L-15 medium with 10% charcoal stripped FBS (Thermo Fisher Scientific, Massachusetts, United States) and 0.04% Ethanol. Culture media in treatment wells were replaced with L-15 medium with 10% charcoal stripped FBS (Thermo

1495 Fisher Scientific, Massachusetts, United States) and β -Estradiol (1 nM, 10 nM
1496 and 100 nM) for estrogen challenge. Each condition was repeated three times.

1497 **3.8.9 Luciferase assay**

1498 At 24 hours after estrogen challenge, the cell culture medium was removed from
1499 each well. 50 μ l of PBS and 50 μ l of DUAL-Glo™ Luciferase Assay buffer
1500 containing DUAL-Glo™ Luciferase Assay Substrate (lyophilized) were added
1501 into each well. Samples were mixed by pipetting 10 times and incubated at room
1502 temperature for 10 min. Samples were transferred to a white walled 96-well flat
1503 bottom plate and scanned for luminescence for 2 seconds per well using a Tecan
1504 Spark plate reader. After repeating the luminescence scan for 3 times per sample,
1505 50 μ l of Dual-Glo® Stop & Glo® was added into each well and mixed by
1506 pipetting 10 times and incubated at room temperature for 10 min. The 96-well
1507 plate containing the mixture was placed back into Tecan Spark plate reader and
1508 scanned for luminescence for 2 seconds per well for Renilla expression as
1509 background control. Luciferase luminescence per well was normalised against
1510 Renilla expression and statistical analysis for normalised luciferase luminescence
1511 was conducted in GraphPad Prism 8 (GraphPad Software, San Diego, USA).

1512 **3.9 Candidate gene functional analysis *in vivo***

1513 **3.9.1 Fish husbandry**

1514 Zebrafish used in this study were raised in recirculatory tank system in the
1515 Temasek Life Sciences Laboratory zebrafish facilities in 28°C water. Fish were
1516 fed twice daily with dried fish pellets (Hikari, Hyogo, Japan) and once daily with

live brine shrimp. Fish used in this study were AB strain zebrafish maintained in the zebrafish facilities in the Temasek Life Sciences Laboratory for more than eight generations.

3.9.2 Guide RNA design and *in vitro* synthesis

To develop sgRNA targeting candidate *POMC* gene in zebrafish, I designed three targets using CRISPR RGEN Tool Cas-Designer^{158,159}. FASTA sequences of each exon were uploaded and designed for SpCas9 protein with the PAM motifs of 5'-NGG-3'. Oligos were ordered for each target specific regions. To synthesize template for in-vitro guide RNA synthesis, I used the oligos listed in Table 5.

Table 5. Oligos used in generating the template for T7 *in vitro* transcription of sgRNA for microinjection

Primer name	Primer sequence (5' – 3')
pomca_t1_danio	TAATACGACTCACTATAGGCTGGCTGTTCTCTGCGCAGGGTTT TAGAGCTAGAAATAGC
pomca_t2_danio	TAATACGACTCACTATAGGGGCTTTCTCCAGGGTAGACGGTTT TAGAGCTAGAAATAGC
pomca_t3_danio	TAATACGACTCACTATAGGCCGCCGGCTAGCAAGCGCTAGTT TTAGAGCTAGAAATAGC
REV_universal	AAAAGCACCGACTCGGTGCCACTTTTTCAAGTTGATAACGGA CTAGCCTTATTTAACTTGCTATTCTAGCTCTAAAAC

Templates for guide RNA synthesis were prepared using a protocol modified from a cloning-free method¹⁵⁰. High fidelity PCR reaction mix was assembled on ice with the following composition: 34 µl PCR-grade water, 1 U of KAPA HiFi polymerase, 1.5 µl of 10 mM target specific primer, 2.5 µl of 10mM REV_universal primer, 1 µl of 10 mM dNTP and 10 µl of 5x KAPA HiFi Buffer. PCR was conducted on a PCR machine with the following run cycles: initial denaturing at 98°C for 30 seconds, 15 cycles of denaturing at 98°C for 10 seconds, annealing at 62°C for 20 seconds, extension at 72°C for 20 seconds, 25 cycles of

1537 denaturing at 98°C for 10 seconds and extension for 72°C for 30 seconds,
1538 followed by a final extension of 72°C for 5 min and hold at 4°C.

1539 Five µl of PCR products were loaded on a 2% agarose gel stained with
1540 0.001% ethidium bromide to check for successful amplifications. Successful
1541 PCR amplification band should be in the 122 bp size range. PCR products were
1542 cleaned with QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). 225 µl
1543 of binding buffer (Buffer PB) was mixed with 45 µl of PCR products by pipetting
1544 gently. The reaction mixture was loaded onto a spin column (QIAquick column)
1545 and centrifuged at 17,900 g for 60 seconds, with flow through discarded. 750 µl
1546 wash buffer (Buffer PE) was added into each column and centrifuged at 17,900
1547 g for 60 seconds, with flow through discarded. Wash step was repeated again and
1548 left to air dry at room temperature for 5 min. 30 µl of RNase-free water were
1549 added into the silica column and incubated at room temperature for 2 min. Elution
1550 was carried out by centrifuging column at 17,900 g for 5 min and eluents were
1551 used for *In vitro* transcription.

1552 sgRNA were synthesized *in vitro* using AmpliScribe™ T7-Flash™
1553 Transcription Kit (Lucigen, Wisconsin, United States). Prior to assembling the
1554 reagents together, all reagents except enzymes were placed at room temperature
1555 for 20 min to prevent the formation of insoluble precipitates. Reagents were
1556 assembled according to the following composition: 3.3 µl of RNase-free water, 3
1557 µl of the template PCR product, 7.2 µl of 25 mM NTP, 2 µl of 10x transcription
1558 buffer, 2 µl of 100 mM DTT, 0.5 µl RNase Inhibitor and 2 µl of T7 RNA
1559 polymerase. Samples were mixed with pipetting up and down for 10 times and
1560 briefly centrifuged before being incubated at 42°C for 4 hours with a heated lid

1561 in a PCR machine. DNase treatment was conducted by adding 1 µl of DNase I
1562 into each reaction mix and incubated at 37°C for 15 min. To purify the transcribed
1563 RNA, DNase digested samples were diluted with RNase-free water to bring the
1564 total volume up to 100 µl. The cleanup was performed using the RNeasy Mini
1565 Kit from Qiagen (Qiagen, Hilden, Germany). 350 µl of silica binding buffer
1566 (Buffer RLT) and 250 µl of absolute ethanol were added to the diluted RNA
1567 samples and mixed by pipetting. Samples were immediately loaded into
1568 QIAGEN silica columns and centrifuged at 8,000 g for 15 seconds. Flow through
1569 was discarded and 500 µl of washing buffer (Buffer RPE) was added into each
1570 column and centrifuged at 8000 g. This washing step was repeated and the
1571 column was centrifuged for 5 min at maximum speed. Any excess wash buffer
1572 found at the edge of the silica column was removed by pipetting and the silica
1573 column was allowed to air dry at room temperature for 10 min. 30 µl of RNase-
1574 free water was loaded into the centre of the silica membrane and incubated at
1575 room temperature for 3 min. Samples were centrifuged at maximum speed to
1576 elute all eluents. The elution was repeated with the eluents from the first round
1577 of elution to increase the recovery of sgRNA. 3 µl of purified sgRNA was loaded
1578 onto 3% agarose gel with 0.001% ethidium bromide staining for electrophoresis
1579 to check for positive transcription. 1 µl of sgRNA was loaded on NanoDrop 2000
1580 (Thermo Fisher Scientific, Massachusetts, United States) to measure for RNA
1581 concentration against a blank reading of RNase-free water. The remaining
1582 samples were aliquoted as 5 µl into multiple small 200 µl single PCR tubes and
1583 stored in -80°C freezers until ready for microinjection.

1584 3.9.3 Microinjection of sgRNA and Cas9 protein

1585 Male and female brooder zebrafish were kept in separate recirculatory tanks until
1586 ready for mating pair set up. Brooder zebrafish were fed at late afternoon the day
1587 before microinjection. On the evening before microinjection, one male zebrafish
1588 were paired up in a breeding tank with two female zebrafish separated by a barrier
1589 and were left overnight. On the next day, after one hour of photoperiod, the
1590 barrier was removed to allow fish mating. Single cell embryos were removed
1591 from the breeding tank and arranged on 2% agarose gel holding plate.
1592 Microinjection mixture was mixed to achieve the following concentration: 250
1593 ng/ μ l of EnGen® Spy Cas9 NLS (New England Biolabs, Massachusetts, United
1594 States) and 200 ng/ μ l of sgRNA in 5 μ l of RNase free water. Microinjection
1595 mixture was incubated in room temperature for 15 min to allow the formation of
1596 ribonucleoprotein (RNP) complex before microinjection. Each embryo was
1597 injected with 1 nl of the mixture at the single cell stage at the animal pole. Injected
1598 embryos were kept in a 28°C incubator and transferred to zebrafish facilities at 5
1599 days post hatch.

1600

1601 3.9.4 Genotyping of *pomc* mutants

1602 At 30 dph, zebrafish were anaesthetized with MS-222. Using a Pasteur pipette,
1603 drop by drop of MS-222 stock solution (10 g/L) were added into the holding tank
1604 until fish showed no swimming movement. Small fin clips were clipped from the
1605 caudal fin located on the tail with small scissors. Scissors were cleaned with 70%
1606 ethanol and burnt on an ethanol lamp in between sampling to prevent cross

1607 contamination. Fish were allowed to recover in individual tanks for two days or
1608 until T7 genotyping assay was completed. DNA extraction from the fin clip was
1609 performed with 100 µl digestion solution containing 10% Chelex (Bio-Rad
1610 Laboratories, California, United States) and 5 µg of proteinase K (Roche Life
1611 Sciences, Basel, Switzerland). Samples were incubated at 55°C for 1 hour and
1612 proteinase K was heat inactivated at 95°C for 5 min. Samples were briefly
1613 centrifuged to pellet the Chelex resins (Bio-Rad Laboratories, California, United
1614 States) and aqueous solution was used for PCR. Primers flanking 600 bp around
1615 target sites were designed with Primer3. Genotyping PCRs were performed with
1616 the flanking primers shown in Table 6. PCR products were reannealed with the
1617 following steps on a PCR machine with heated lid: 5 min denaturing step at 95°C
1618 and dropped to 25°C at a ramping rate of -0.1°C/sec. Reannealed PCR products
1619 were digested with T7 endonuclease I at 37°C for 60 min. Digestion was stopped
1620 with 25 mM EDTA and products were resolved with 2.5% agarose gel
1621 electrophoresis and imaged with 0.001% ethidium bromide staining.

1622 Table 6. Primers flanking target sites of POMC sgRNA for genotyping assay and
1623 qPCR

Primer Name	Primer Sequence
pomca_t1_genof	TGGGTGAAAAATATCCCATACACT
pomca_t1_genor	TGATAGTTGCCCAACAGACAT
pomca_t2_genof	AAAAAGTTTAACAAAACATTTTGC GA
pomca_t2_genor	CTCCAGCGGAAATGGGTCAT
pomca_t3_genof	CCTTCCCCTGTAGCACTCG
pomca_t3_genor	CTGTTTTAGCCTCTAAGCTGGT
pomc-genof	GTTTGGTGACCTTGAGGAATTGACG
pomc-genor	TTGGCAACCGCTAAGTGATTTTGAG
pomc-q3F	TGTCGAGACCTCAGCACAGA
pomc-q3R	GCTCGTCCCAGGACTTCAT

1624 3.9.5 Sequencing of genome editing sites

1625 PCR products from section 4.9.4. were cleaned with QIAquick PCR Purification
1626 Kit (Qiagen, Hilden, Germany) and cloned into Promega pGEM®-T Vector
1627 Systems (Promega, Fitchburg, USA). Single colony clones were obtained
1628 through blue-white screening. White colonies were sub-cultured in LB medium
1629 containing Ampicillin and used for colony PCR using M13 sequencing primers.
1630 M13F were used for BigDye sequencing. Sequences were cleaned with
1631 Sequencher V5.1 (Gene Codes Corporation, Michigan, United States) and
1632 sequence alignment was performed using Muscle in MEGA software¹⁹⁵.

1633 To check for large deletions, I designed a pair of primers flanking POMCa
1634 gene in zebrafish, with a target size of 3 kb and T_m of 61°C. PCRs were
1635 conducted with the following run cycles: initial denaturing at 98°C for 30 seconds,
1636 30 cycles of denaturing at 98°C for 30 seconds, annealing at 58°C for 30 seconds,
1637 extension at 72°C for 120 seconds, followed by a final extension of 72°C for 10
1638 min and hold at 4°C. PCR products were checked for double bands on 1% agarose
1639 gel stained with 0.001% ethidium bromide.

1640 To investigate the *pomc* expression level of *pomc* mutants, four *pomc*
1641 mutants and four wildtype zebrafish were sampled at 10 dph. RNA samples were
1642 extracted using Favorgen FavorPrep Tissue Total RNA Mini Kit. Whole
1643 embryos were homogenized with a handheld electric homogenizer in 350 µl of
1644 FARB Buffer and 3.5 µl of β-Mercaptoethanol and incubated at room
1645 temperature for five minutes. Homogenized samples were filtered on a spin
1646 column using microcentrifuge at 18,000 x g. The clarified supernatant was then
1647 mixed with one volume of 70 % RNase-free ethanol by vortexing. Samples were

transferred into a FARB mini column and centrifuged at 18,000 x g for one minute. Five hundred µl Wash Buffer 1 were added into the silica tubes and centrifuged at full speed for one minute. Then, 750 µl of Wash Buffer 2 were added into the silica column and centrifuged at 18,000 g for one minute and the supernatant was removed. The silica tubes were centrifuged again at full speed for five minutes to remove all excess alcohol. RNA samples were eluted using 50 µl of RNase-free water. Reverse transcription was carried out using M-MLV reverse transcriptase with a mix of random hexamer primers and oligo-dT primers. cDNA products were diluted 10-fold with water prior to qPCR reaction. Real time PCR was conducted using KAPA SYBR Fast Universal Kit (Sigma-Aldrich, Missouri, United States) in CFX96 Touch™ Real-Time PCR Detection System (Bio-Rad Laboratories, California, United States). qPCR data analysis was conducted using the delta delta Ct method. Primer sequences for qPCR analysis can be found in Table 6.

3.9.6 Phenotypic observation of *pomc* mutants

3.9.6.1 Melanophores pigmentation phenotyping

To observe the phenotypic difference between *pomc* mutants and wildtype zebrafish, fish were reared to 100 dph and anaesthetized by adding drop by drop of MS-222 (10 g/L) until fish showed “belly up” behaviour. Fish were placed on a petri dish and observed under a Leica MZ FLIII (Leica, Wetzlar, Germany) stereomicroscope. Images were taken with an attached Nikon DXM1200F (Nikon, Tokyo, Japan) and imaging software ACT-1 V2.7 (Nikon, Tokyo, Japan). Fish were imaged, in RGB TIFF format, from the operculum to the mid-trunk on the lateral view and from the mid-trunk to the tail. Images were loaded into R-

environment and pre-processed with the command `processimages(jpglist, scalefactor=0.1, threshold=0.3)`, which preprocessed the melanophores as black pixels using the library package `imageR V0.41.2`. The proportion of black pixels was calculated as the ratio of black pixels to the total number of pixels in the whole image. Statistical test to test the differences in pigmentations was done using the `t.test` function in R.

3.9.6.2 Bodyweight and body length phenotyping of *pomc* mutants

To test if *POMC* knock-out mutation will affect growth rates, I crossed two heterozygous *pomc* F0 founders and raised the embryos up to 80 dph in a recirculatory tank. Fish were fed twice daily with dry fish pellet (Hikari, Hyogo, Japan) and once daily with live brine shrimp. At 80 dph, fish were anaesthetized by adding drop by drop of MS-222 (10 g/L) until fish showed “belly up” behaviour. Fish were gently patted dry with a paper towel and weighed on a PR2003 Delta Range (Mettler Toledo, Ohio, United States) digital weighing scale. Each fish was then placed on a transparent petri dish and its total length (mouth to tail) measured with a ruler beneath the petri dish. Using a scalpel, the caudal fins of the embryos were sampled and used for DNA extraction and T7 genotyping as described in section 3.9.4. Statistical analysis to compare mutants and wildtype zebrafish were performed using Estimation Stats two groups student’s T-test with Hedge’s G option and 95% confidence intervals to correct for small sample size bias (<http://www.estimationstats.com/#/analyze/two-independent-groups>)¹⁹⁶.

1695 **3.9.6.3 Feeding behaviour phenotyping**

1696 To observe the feeding behaviour differences between *pomc* mutants and
1697 wildtype fishes, I crossed a pair of wildtype parents and a *pomc* F0 mosaic parents
1698 and raised the juveniles to 20 dph for feeding behaviour phenotyping. Fish were
1699 fed with mixed juvenile dry feed pellets ground to 5 microns in size, including
1700 commercial feed and algae powder. Fish were fed twice daily with the same
1701 amount of food up to the day of the experiment.

1702 Video taking was started at 10 seconds before the addition of feed and
1703 stopped at 50 seconds after the feed was added. Fish latency to feeding was
1704 calculated by measuring the time taken by the fish from feed addition to fish
1705 rising to the water surface for the first ten fish for feed foraging. Observations
1706 were repeated three times for each group. Videos were taken with a mobile phone
1707 video camera (Huawei, Shenzhen, China) and transferred to a computer for
1708 analysis using VLC video player at 0.25x video playback speed setting. Feeding
1709 latency timing was calculated with the following formula:

1710 Feeding latency (s) = time of Surface to forage (s) – time of Feed addition (s)

1711 Statistical analysis of feeding behaviour was conducted with Estimation Stats two
1712 groups student's T-test with Hedge's G option and 95% confidence intervals to
1713 correct for small sample size bias
1714 (<http://www.estimationstats.com/#/analyze/two-independent-groups>).

1715

1716 **4 Results**

1717

1718 **4.1 Whole genome bisulfite sequencing**

1719 **4.1.1 Single base-pair resolution methylome of tilapia**

1720 A total of 107.6 Gb of sequencing data averaging 26.69 Gb per sample were
1721 obtained from whole genome bisulfite sequencing of tilapia muscle tissue. A
1722 statistical summary of various sequencing result parameters is presented in Table
1723 7. Clean reads filtered and cleaned using NGS QC Toolkits¹⁹⁷ sum up to a total
1724 of 242 million read pairs, averaging ~60 million read pairs per library. A total of
1725 182,639 reads were mapped onto the Lamda DNA reference genome (GenBank
1726 ID: J02459.1), out of which 0.30% of the cytosines were read as C instead of T,
1727 indicating that the bisulfite conversion efficiency is 99.7%. Reads were mapped
1728 to the oreNil2 Nile tilapia reference genome assembly using a three-letter aligner,
1729 *bismark*¹⁹⁸. I used *bowtie2* algorithm in *bismark* as *bowtie2* gives better
1730 performance for reads with read length longer than 50 bp¹⁹⁹. I allowed one
1731 mismatch per read to increase the sensitivity of our mapping process, although
1732 this came at a cost of higher computing power requirement and a much longer
1733 mapping time.

1734

1735 Table 7. Statistical summary of skeletal muscle whole genome bisulfite
1736 sequencing

Library	Male1	Male2	Female1	Female2
# Raw Reads	46738562	59386546	60160140	75836536
# Mapped Reads	28185036	36742798	35630170	44371404
Mapping Efficiency	56.1%	56.7%	54.6%	54.0%
Average Sequencing Coverage	5.61x	7.26x	7.08x	8.80x
Bisulfite Conversion Efficiency (%)	99.75	99.68	99.65	99.72
CpG Sites Covered at 2X Coverage (%)	78.52	83.74	87.87	96.17
CpG Sites Covered at 5X Coverage (%)	29.37	48.32	53.67	60.31
CpG Sites Covered at 10X Coverage (%)	9.53	27.85	26.49	27.39
CpG Methylation Level (%)	69.64	69.67	69.50	69.68
CHG Mehtylation Level (%)	0.49	0.49	0.49	0.43
CHH Methylation Level (%)	0.58	0.59	0.61	0.48

1737

1738 The mapping efficiency for each library is 56.00%, 56.60%, 54.50% and
1739 53.90% for Male-1 (M1), Male-2 (M2), Female-1 (F1) and Female-2 (F2)
1740 libraries respectively. Cytosine methylation levels at single base-pair resolution
1741 were extracted using the software methylKit in R-environment¹⁰⁶. Using the
1742 read.bismark function, a minimum reads coverage of 5x was set as the minimum
1743 parameter for calling CpG, CHH and CHG methylation states from sorted SAM
1744 file outputs generated by bismark¹⁹⁸. An average of 20,334,256 cytosine
1745 methylation positions (the combination of CpG, CHG and CHH at both Watson
1746 and Crick strands) were called for each sample. Based on our mapping results,
1747 on average 69.60% of cytosines in CpG context are methylated. In CHH and
1748 CHG context, only 0.57% and 0.47% of the cytosines are methylated respectively.
1749 In Figure 5, I present a comparison of genome-wide CpG methylation levels
1750 between the four samples studied in this report. Overall, the samples showed

1751 similarities of CpG methylation profile on a genome-wide scale. Based on an
 1752 averaged Pearson correlation of 0.831, DNA CpG methylation profiles are
 1753 largely conserved between male and female tilapias. However, there are distinct
 1754 regions of sexually dimorphic DMRs found in our analysis, which will be
 1755 discussed in subsequent paragraphs.

1756

1757

1758 Figure 5. Correlation matrix showing the Pearson correlation of base resolution
 1759 CpG methylation genome-wide between the skeletal muscle tissue samples of
 1760 hybrid tilapia. Histograms show CpG methylation level of each sample from 0%
 1761 to 100% distributed across 20 bins of 5% intervals. The red line and green line
 1762 represent linear regression and loess fit, respectively, to model the relationship
 1763 of differential CpG methylation sites between compared individual pairs. The

1764 high R^2 values between samples suggest that the majority of CpG sites have
1765 similar DNA methylation level.
1766

1767 **4.1.2 Tilapia gene body methylation patterns**

1768 To investigate the methylation patterns of tilapia around the gene structure, I
1769 extracted the Transcriptional Start Sites (TSS), Transcriptional Termination Sites
1770 (TTS) and mRNA coordinates from the reference annotation. I then combined all
1771 four libraries and proceeded to generate one high coverage methylome ($> 40 \times$
1772 coverage) using the same procedure described above. I extracted the coordinates
1773 of 5 Kb upstream and 5 Kb downstream of the TSS and TTS respectively and
1774 divided them into 100 bins. The average CpG methylation value for each bin was
1775 then called using SeqMonk
1776 (<http://www.bioinformatics.babraham.ac.uk/projects/seqmonk/>). mRNA tracks
1777 of 28,730 genes were downloaded from UCSC Nile tilapia GeneScan tracks and
1778 divided into 100 bins each. The mean methylation of each bin was calculated by
1779 averaging the methylation level of every CpG cytosine within the range. A gene
1780 structure methylation profile is shown in Figure 6. The CpG methylation level of
1781 tilapia, on average, drops to 23.00% approaching the TSS. Along with the gene
1782 bodies, CpG methylation increased rapidly after TSS and reached a plateau of
1783 around 75.00% before dropping to 50.00% approaching the TTS. Between the
1784 TTS and 5 Kb downstream of TTS, CpG methylation levels averaged at 70.00%.
1785

1786

1787 Figure 6. CpG Methylation level in relation to gene body, 5 Kb region upstream
 1788 of TSS and 5 Kb downstream of TTS in hybrid tilapia. CpG methylation level
 1789 dropped gradually to 25% approaching the TSS before rising sharply to 75% in
 1790 the gene body. CpG methylation level then dipped rapidly to 50% approaching
 1791 the TTS and subsequently returned to earlier intragenic CpG methylation level.

1792

1793 4.1.3 Methylation profiles of various repeat elements in tilapia reference 1794 genome

1795 I annotated the tilapia reference genome for repeat elements using RepBase (26-
 1796 5-2016) and RepeatMasker¹⁷⁸. The parameters used were: **-s -gff -species**
 1797 **vertebrates**. I selected **-s** (slow option) as it provides up to 5% more sensitivity
 1798 but is 3 times slower than the default parameter as described by RepeatMasker
 1799 authors¹⁷⁸. I also set the RepeatMasker search mechanisms to target only repeats
 1800 found in vertebrates to reduce the numbers of false positives. In this annotation,
 1801 I masked approximately 153 Mb (16.51% of the entire genome) as repeat
 1802 elements. The annotation is summarized in Table 8. Up to 4.78% of the genome
 1803 was found to be retro elements such as short interspersed nuclear elements

1804 (SINEs), long interspersed nuclear elements (LINEs) and long terminal repeats
1805 (LTRs). I also annotated 3.58% of the reference genome as DNA transposons.
1806 1.35% of the genome was found to be simple repeats. I then calculated the total
1807 length of transposable elements in each chromosome divided by the length of
1808 each chromosome. I found no chromosome with an outlying proportion of TEs,
1809 after calculating the Z-score of TE percentage in each chromosome.
1810

1811 Table 8. Summary of repeat elements masked with RepeatMasker based on
 1812 vertebrate repeat element homology search in the tilapia genome

Name of Elements	Number of Elements	Length (kb)	Percentage in Genome (%)
Retroelements	206987	53298.94	5.75
SINEs:	45918	6640.64	0.72
Penelope	19438	3546.54	0.38
LINEs:	137810	38082.12	4.11
CRE/SLACS	0	0.00	0.00
L2/CR1/Rex	67908	18088.45	1.95
R1/LOA/Jockey	1600	544.69	0.06
R2/R4/NeSL	7814	3209.57	0.35
RTE/Bov-B	27784	7239.29	0.78
L1/CIN4	10127	4671.61	0.50
LTR elements:	23259	8576.17	0.92
BEL/Pao	2056	1697.22	0.18
Ty1/Copia	241	210.80	0.02
Gypsy/DIRS1	9455	5445.19	0.59
Retroviral	8414	953.74	0.10
DNA transposons:	245024	42830.38	4.62
hobo-Activator	69080	7550.16	0.81
Tc1-IS630-Pogo	105934	27576.22	2.97
En-Spm	0	0.00	0.00
MuDR-IS905	0	0.00	0.00
PiggyBac	4845	794.34	0.09
Tourist/Harbinger	5538	860.73	0.09
Other (Mirage, P-element, Transib)	818	81.81	0.01
Total interspersed repeats		97963.14	10.56
Small RNA:	2256	193.27	0.02
Satellites:	3332	524.20	0.06
Simple repeats:	318248	12238.05	1.32
Low complexity repeats:	49226	2344.19	0.25

1813

1814

1815 I profiled the CpG methylation levels of various genomic elements as high (> 75%
1816 CpG methylation), medium (75%-25% CpG methylation) and low (< 25% CpG
1817 methylation) for both male and female skeletal muscle tissue. I found that in both
1818 sexes, more than 50% of DNA repeat elements such as DNA transposons, LINEs,
1819 SINEs, LTRs and satellites are highly methylated (> 75% CpG methylation). In
1820 promoter regions, however, close to 40% of the CpG sites are lowly methylated
1821 (< 25% CpG methylation). Only 25% of the promoter regions are highly
1822 methylated in both sexes.
1823

1824

1825 Figure 7. Distribution of CpG methylation level across various genomic elements
 1826 such as promoters, introns, exons, DNA transposons, SINEs, LINEs and LTRs
 1827 in male and female tilapia.

1828

1829 4.1.4 Sexually dimorphic differentially methylated regions revealed by 1830 sliding windows analysis.

1831 Base-pair resolution of methylome was extracted from bismark output SAM files
 1832 using the R-package methylKit¹⁰⁶. I imposed a minimum requirement of 5x

1833 coverage per cytosine in order to call on the CpG methylation percentage for each
1834 CpG position. Our data shows that at 5x coverage, 47.9% of all CpG sites were
1835 covered with minimally 5 reads on average. Samples were divided into biological
1836 replicate sets of male and female with two replicates each. Potential PCR
1837 replicates were removed by filtering out the top 0.1% CpG with the highest
1838 sequencing coverage. While the CpG methylation profiles were largely similar
1839 between the samples, I found regions of significant DMRs between the male and
1840 female tilapia at both base and region levels. For CpG base-level differential
1841 methylation analysis, 361,702 genome-wide CpG coordinates with a minimum
1842 5x coverage in every sample were extracted for differential methylation
1843 calculation. A total of 1,717 CpG sites were found to be differentially methylated
1844 between male and female tilapia muscle tissue with a q-value (FDR corrected p-
1845 value) cut-off point of 0.05. However, single base-pair methylation is often not
1846 informative enough to infer functions for DMRs on a genomic scale, partly due
1847 to missing data points which gave heavy penalties to the q-values when
1848 calculating differential methylation. A sliding window analysis is more
1849 appropriate as it represents regions of differentially methylated cytosines instead
1850 of single cytosines¹⁰⁶. Also, the methylation level of each window can be
1851 averaged across a designated window frame, providing compensation for lost
1852 single base-pair data points or CpG sites with low coverage in some libraries.
1853 Only genomic sequences anchored onto chromosomes were used in this analysis.
1854 Contigs or scaffolds were exempted from downstream analysis. Using the
1855 calculateDiffMeth function in methylKit¹⁰⁶, a total of 4,757 windows were
1856 identified as DMRs with a q-value (FDR corrected p-value) cut-off point of 0.01
1857 and minimal CpG methylation difference of 25%. The DMRs have a combined

1858 size of 4.76 Mb, representing 0.45 % of the genome assuming the genome size
1859 of tilapia is 1060 Mb. The CpG methylation landscape in the two sexes of tilapia
1860 is summarized in Figure 8 in CIRCOS format, shown along with the density of
1861 CpG islands in each chromosome.

1862

1864

1865 Figure 8. CIRCOS chart summarizing the differentially methylated regions in
1866 two sexes of hybrid tilapia. Track 1 shows an ideogram representing the 22
1867 chromosomes in hybrid tilapia. Track 2 represents differentially methylated
1868 regions between male and female individuals. Histograms in the blue regions
1869 show regions hypermethylated in males while orange regions show regions
1870 hypermethylated in females. Histograms are expanded for illustration purposes
1871 and histogram width is not a direct representation of DMR size. Track 3 displays
1872 the locations of all CpG islands in hybrid tilapia.

1873

1874 The identified DMRs with more than 25% difference in hypermethylated
1875 CpG levels are visualized in Figure 9. In this graph, high numbers of DMRs
1876 located on LG1 can be observed based on a Z-score of 3. Besides LG1,
1877 moderately high numbers of DMRs were also observed to be located on LG2,
1878 LG7, LG16-21, LG18 and LG23. Some of these linkage groups were previously
1879 shown to be associated with sex determination.

1880

1881

1882 Figure 9. Numbers of hypermethylated DMRs in male (blue) and female (red)
1883 tilapia skeletal muscle tissue. The dotted line graph represents the Z-score of
1884 number of DMRs in each chromosome. In this plot, chromosome LG1 shows a
1885 very high number of DMRs as compared to the other chromosomes (Z-score > 3).

1886

1887 **4.1.5 CpG methylation in promoters is negatively correlated with the gene** 1888 **expression**

1889 To investigate a possible relationship between gene promoter CpG methylation
1890 level and gene expression level, I extracted the CpG methylation level for an
1891 arbitrary promoter region 1000 bp upstream from transcriptional start sites for
1892 each gene and compared the methylation level to the respective mRNA
1893 expression level. I found a general trend whereby low levels of CpG methylation
1894 in the arbitrary promoter regions were correlated with increased gene expression
1895 level. On the other hand, high CpG methylation in the arbitrary regions is
1896 correlated with low to silent gene expression. A scatterplot with a distinct L-
1897 shaped pattern was produced when the absolute methylated CpG read count in
1898 the arbitrary promoter region was plotted against gene expression level (RPKM)
1899 for each corresponding gene as shown in Figure 10. Gene suppression effect was
1900 most robust when the number of methylated cytosine in the corresponding
1901 promoter region was high. Our data extend the conventional view that promoter
1902 CpG methylation in teleost genomes can suppress gene expression, similar to the
1903 observations found in mammalian genomes^{101,102,200-202}.

1904
 1905 Figure 10. Scatterplot showing relationship between numbers of methylated CpG
 1906 and gene expression levels (RPKM). In promoters with more methylated CpG
 1907 sites, gene expressions are suppressed. Legend on the right shows colouring
 1908 schemes for absolute numbers of methylated CpG reads (mC_reads).

1909

1910 4.1.6 Whole genome bisulfite sequencing reveals sexually dimorphic 1911 regulatory landscape

1912 I intended to predict and infer the locations of regulatory regions of hybrid tilapia
 1913 with the aim of identifying cis regulatory elements that may provide evidence to
 1914 explain the phenomenon of sexual size dimorphism. Using the WGBS data
 1915 generated for tilapia, I inferred the active cis-regulatory regions (ACRs) from the

1916 tilapia genome using the R software MethylSeekR. I generated a Biostrings-
1917 based R library, BSgenome.Oniloticus.UCSC.orenil2, as the reference package
1918 for MethylSeekR analysis. A methylation cut-off point of 50% and a false
1919 discovery rate threshold of 0.05% were set as parameters for segmentation of
1920 genomic regions into LMRs and UMRs, as recommended by the authors of
1921 MethylSeekR¹⁰³. Partially methylated regions (PMDs) were identified and
1922 masked. PMDs were defined as contiguous regions with an average methylation
1923 level of less than 70%. These PMDs are characterized by highly disordered
1924 methylation, resulting in average methylation clearly below the genomic
1925 background level. It is essential to mask these PMDs to reduce false positive
1926 hits¹⁰³. From a previous study done in mice, low methylated regions (LMRs)
1927 were found to be associated with active occupancy by DNA-binding factors,
1928 suggesting an active cis-regulatory region or active transcription factor binding
1929 sites¹⁰³. Using the algorithms in MethylSeekR, I intended to infer the ACRs in
1930 the tilapia genome with the aim to uncover regions of transcription factor
1931 occupancies. From the WGBS datasets, I inferred 101,476 LMRs and 16,154
1932 UMRs from the female methylome while 83,697 LMRs and 16,318 UMRs were
1933 inferred from the male methylome. A sample region illustrating LMRs and
1934 UMRs can be found in Figure 11. I then defined arbitrary promoter regions as
1935 500 bp upstream of transcriptional start sites and performed enrichment
1936 quantifications using SeqMonk for LMRs that are present within these arbitrary
1937 promoter regions. The top 50 promoter regions for each sex with the highest
1938 enrichment for LMRs were selected to extract the FASTA sequence and were
1939 analyzed using the MEME suite for motif analyses¹⁷⁹.

1940

1941

1942 Figure 11. Representative illustration of identified active cis-regulatory region in the tilapia genome using whole genome bisulfite sequencing.
1943 Coordinates of masked PMRs (green line), LMRs (red triangle) and UMRs (blue square) in a portion of chromosome LG15 are illustrated here.
1944 Each black dot represents a CpG coordinate (X-axis) and its CpG methylation level (Y-axis). The top row represents raw data while the bottom
1945 row represents window-averaged data.

1946

1947 I selected five common motifs returned with the lowest E-value and
1948 proceeded to search for their corresponding transcription factors from the
1949 JASPAR database²⁰³. The resulting motifs and the associated transcription
1950 factors are listed in Figure 12. These transcription factors are responsible for the
1951 development of skeletal and heart muscle (RUNX2, FOXP1, MTF1, and FOXJ3)
1952 and the development of the nervous system (EGR2, EGR4, FOXC1 and
1953 TFAP2E). I also found increased enrichment of transcription factor binding site
1954 motifs that are associated with the immune system in the male group (ETS1, SPI1,
1955 PRDM1, ELF1 and IRF1). In the female group, I found an enrichment of motifs
1956 related to the cell cycle (E2F3, E2F4, and E2F6), development (EGR2b, TFAP2e
1957 and FOXC1a) and negative regulation of epithelial cells (EHF).
1958

1959

Figure 12. Enrichments of sexually dimorphic active cis-regulatory motifs in hybrid tilapia skeletal muscles. Consensus sequence motifs and their associated transcription factors on promoters enriched with lowly methylated regions (LMRs) in females and males. Several reports have shown that LMRs are bound with transcription factors, which suggests active transcription activities^{103,204}.

1960

E-value	Frequency	Associated Transcription Factors
7.4e ⁻³⁹	36%	EGR2, RREB1, GM397, RUNX2, EGR4
2.2e ⁻²⁴	66%	ZNF263, ETS1, SPI1, PRDM1, ELF1
1.0e ⁻¹⁵	80%	FOXP1, MTF1, ZFP105, ONECUT3, IRF1
1.6e ⁻¹²	26%	No Hits
4.5e ⁻⁹	58%	ZFP281, SP1, ZNF263, EGR1, KLF5

1967 4.2 Muscle transcriptome

1968 4.2.1 Identification of sex-biased genes in tilapia muscle

1969 The Nile tilapia reference genome assembly used for RNA-seq analysis is the
1970 O_niloticus_UMD_NMBU assembly. The percentage of uniquely mapped reads
1971 for M1, M2, F1 and F2 was 92.33%, 92.93%, 92.91% and 92.87%, respectively.
1972 Differentially expressed genes (DEGs) were defined as having a log2 fold-
1973 change of more than 1 and an adjusted *p-value* of less than 0.05. In our analysis,
1974 I found 136 sexually dimorphic DEGs in the skeletal muscle transcriptome.
1975 Twenty-three genes were highly expressed in the male tilapia transcriptome,
1976 while 113 genes were highly expressed in the female tilapia transcriptome.
1977 Among the 23 male-biased genes, one gene belongs to the non-coding RNA class
1978 while the rest are coding RNA. Out of the female-biased genes, four genes belong
1979 to the non-coding RNA class while the remaining genes are protein coding RNAs.
1980 From our list of sex-biased genes, I identified five candidate genes that are related
1981 to somatic growth and regulation. These genes are *RASGRF1*, *LPIN1*, *FOXO4*,
1982 *IRS2* and *EGFR*.

1983 Table 9. Summary of RNA-seq statistics for skeletal muscle transcriptome

Library	# Raw Reads	# Mapped Reads	Mapping Efficiency (%)
Male1	19230693	17,755,699	92.33
Male2	20922517	19,443,295	92.93
Female1	18366944	17,064,728	92.91
Female2	17932753	16,654,148	92.87

1984

1985 **4.2.2 Validation of RNA-seq data with qRT-PCR**

1986 To validate our RNA-seq data, I randomly selected six genes from the list of sex-
1987 biased genes identified, and designed primers at the 3' end of the mRNA
1988 sequences. Gene expression data were normalized using actin as housekeeping
1989 gene control. Log₂ fold-change result for each gene obtained from RNA-seq
1990 quantification and qPCR quantification were compared. An R-squared value of
1991 0.8312 were obtained for the muscle transcriptome comparison. The linear
1992 correlation curve was significantly non-zero at $p = 0.01$. The correlation linear
1993 curve can be found in Supplementary Figure 1.

1994

1995 **4.2.3 Chromosomal location of skeletal muscle sex-biased genes.**

1996 Based on the latest genome assembly, O_niloticus_UMD_NMBU, I performed a
1997 Chi-square analysis to test if the sex-biased genes identified are randomly
1998 distributed across various chromosomes in the tilapia genome. If the genes are
1999 randomly distributed across different chromosomes, the number of sex-biased
2000 genes in each linkage group should be proportional to the number of genes within
2001 the chromosome. By comparing the observed and expected numbers of sex-
2002 biased genes for each chromosome, I rejected the null hypothesis that the sex-
2003 biased genes are randomly distributed across the genome (Chi-squared test $P <$
2004 0.001). After calculating the Z-score of sex-biased gene count for each
2005 chromosome, I found that chromosomes LG4 and LG18 have Z-scores of more
2006 than 1.645 (significant at $P < 0.05$), suggesting that sex-biased gene expression
2007 in the muscle tissues are enriched in these chromosomes as shown by Figure 13.

2008 Besides the non-random distribution of the sex-biased genes, three genes were
2009 also found to be located within a QTL for growth previously identified in tilapia
2010 in chromosome LG18⁴⁴. These genes are LOC100706127, CADM3 and a non-
2011 coding RNA (LOC109195677). The functions of LOC100706127 are unknown.
2012 A protein BLAST search showed the presence of an immunoglobulin domain and
2013 similarity with proteins from signalling lymphocytic activation molecule (SLAM)
2014 family, suggesting a role in immune functions. CADM3 functions as a cell to cell
2015 adhesion protein²⁰⁵. The expression of CADM3, also known as NECL1, can
2016 suppress the growth and tumorigenic ability of colon cancer cells²⁰⁵. The role of
2017 CADM3 in skeletal muscle growth is still unknown. LOC109195677 is a novel
2018 predicted non-coding RNA. A reciprocal BLAST search with LOC109195677
2019 sequence to detect orthologues revealed no hits.

2020

2021

2022

2023 Figure 13. Numbers of sex-biased genes on different chromosomes are shown on
2024 the bar plot. Z-scores were calculated and shown on the right Y-axis with the line
2025 graph. Chi-square analysis showed that the sex-biased genes are not randomly
2026 distributed across the genome ($P < 0.001$).

2027

2028 **4.2.4 Identification of zebrafish orthologues using BLAST.**

2029 From the O_niloticus_UMD_NMBU assembly, there are 62032 coding
2030 sequences available, which includes various isoforms and transcript variants. All
2031 coding sequences were searched for homology against 52089 coding sequences
2032 in the zebrafish genome (GRCz11) using BLAST. I obtained 44281 1:1 matches
2033 for tilapia:zebrafish orthologues. After removing redundancies from genes with
2034 multiple spliced isoforms, I obtained 20467 ortholog gene pairs. These zebrafish
2035 ortholog hits were used to match gene ontologies information for tilapia genes.

2036

2037 **4.2.5 PantherGO and KEGG pathway mapping.**

2038 Zebrafish gene ID orthologous to the sex-biased genes in tilapia were uploaded
2039 into PANTHER DB (<http://www.pantherdb.org/geneListAnalysis.do>) for
2040 statistical overrepresentation test²⁰⁶. I applied Fisher's exact test, with a false
2041 discovery rate (FDR) calculation using a cut-off of $P < 0.05$. From the 126
2042 orthologues of sex-biased genes in skeletal muscle, I observed overrepresentation
2043 in genes relating to various biological processes including muscle system process
2044 (GO:0003012), anatomical structure morphogenesis (GO:0009653), actin
2045 filament-based process (GO:0030029), multicellular organismal process
2046 (GO:0032501), actin cytoskeleton organization (GO:0030036), regulation of
2047 muscle contraction (GO:0006937), regulation of muscle system process
2048 (GO:0090257), circadian regulation of gene expression (GO:0032922) and
2049 muscle contraction (GO:0006936). On the cellular compartment GO segment,
2050 troponin complex (GO:0005861), myofilament (GO:0036379), striated muscle

2051 thin filament (GO:0005865), sarcomere (GO:0030017), contractile fiber part
 2052 (GO:0044449), myofibril (GO:0030016), contractile fiber (GO:0043292), I band
 2053 (GO:0031674), actin cytoskeleton (GO:0015629), supramolecular fiber
 2054 (GO:0099512), supramolecular polymer (GO:0099081), supramolecular
 2055 complex (GO:0099080) and cytoskeleton (GO:0005856) were overrepresented.
 2056 All GO enrichment tests were conducted using Fisher's exact test, with the
 2057 following significance criteria: false discovery rate corrected p-value score of
 2058 less than 0.05 and enrichment fold of more than 2. Gene ontologies of sex-biased
 2059 genes are summarized in Figure 14 generated with R software ClusterProfiler¹⁸³.

2060
 2061 Figure 14. ClusterProfiler diagram summarizing the overrepresentations of gene
 2062 ontologies in sex-biased genes identified in tilapia skeletal muscle.

2063 **4.2.6 Female sex-biased genes exhibit higher evolution rates.**

2064 Some sex-biased genes in tilapia skeletal muscles are evolving at a faster rate
 2065 than non-biased genes (Mann-Whitney U Test; $P < 0.001$), having a higher dN/dS
 2066 ratio as shown in Figure 15. When I compared female-biased genes against non-
 2067 biased genes, I observed significantly higher non-synonymous to synonymous

2068 ratios (dN/dS) in female-biased genes (Mann-Whitney U Test; $P < 0.001$).
2069 However, when I compared the male-biased genes against non-biased genes, I
2070 observed no significant differences (Mann-Whitney U Test; $P = 0.138$). I also
2071 compared the median dN/dS values for female-biased and male-biased genes
2072 against unbiased genes and found that they are 17.4% and 13.5% higher than
2073 unbiased genes, respectively.

2074 I found a dN difference of 0.0052 between female-biased and unbiased
2075 genes (Mann-Whitney U Test; $P < 0.001$) but no significant difference in dN
2076 between male-biased genes and unbiased genes (Difference of 0.0019, Mann-
2077 Whitney U Test; $P = 0.026$). The difference in dS between male-biased and
2078 female-biased genes against un-biased were insignificant. Hence, the elevated
2079 dN/dS values in female-biased genes were due to the higher increase in non-
2080 synonymous substitution rate (dN) and not due to the decrease of synonymous
2081 substitution rate (dS).

2082

2083

2084

2085 Figure 15. Non-synonymous to synonymous mutation rate ratio (dN/dS) of
2086 female-biased (N = 366), male-biased (N = 92) and randomly selected unbiased
2087 genes (N = 145). (***) $P < 0.01$

2088

2089 As positive selection signals may be masked by genes with longer coding
2090 sequences, I used $dN/dS > 0.5$ as a threshold for positive selections, as used by
2091 others^{99,207}. Using this cut-off value, I identified several candidate genes that are
2092 suggestive of a positive selection event. These genes are: *PODXL* ($dN/dS =$
2093 1.374), *MYH7B* ($dN/dS = 0.956$), *SLC26A10* ($dN/dS = 0.693$), *TNNT1* ($dN/dS =$
2094 0.662), *RAPGEF5a* ($dN/dS = 0.580$), *MH7l* ($dN/dS = 0.574$) and *CLK4a* (dN/dS
2095 = 0.508). More than 76%, 83% and 85% of the male-biased, female-biased and
2096 unbiased genes, respectively, have higher codon adaptation index than reference
2097 codon adaptation index. There were no significant differences in codon
2098 adaptation index between groups. I then extracted the gene expression level for
2099 each gene in sex-biased groups and performed a correlation analysis between
2100 gene expression level and the codon adaptation index. I found a positive

2101 correlation ($R = 0.396$, $P < 0.01$) between gene expression level and codon
2102 adaptation index. This is consistent with studies done in *Drosophila*
2103 *melanogaster* and *Ectocarpus spp.*, where the codon adaptation index correlates
2104 strongly with high gene expression levels^{114,208}. When I compared the codon
2105 adaptation index between male and female-biased genes, no significant
2106 differences between these two groups were found.

2107

2108 **4.2.7 Identification of SNPs in RASGRF1 promoter region and its** 2109 **association with bodyweight in tilapia.**

2110 To genotype and discover potential markers that may be applied in marker
2111 assisted selection for growth, I looked for SNPs close to the proximity of
2112 candidate genes. I identified three SNPs located in the promoter region of
2113 RASGRF1. The coordinates of these SNPs are LG1:28146922 (C/T),
2114 LG1:28146925 (C/T) and LG1:28146966 (A/G), based on the genome assembly
2115 O_niloticus_UMD_NMBU. The location of the SNPs LG1:28146922 (C/T),
2116 LG1:28146925 (C/T) and LG1:28146966 (A/G) are located at 877 bp, 880 bp
2117 and 921 bp upstream of the transcriptional start site of RASGRF1, respectively.
2118 In our genotyping population of 158 samples, I did not find any recombination
2119 between these three SNPs locus. From our population, individuals with the
2120 genotypes LG1:28146922 (CT), LG1:28146925 (CT) and LG1:28146966 (AG)
2121 weigh on average 350.9g while individuals with the genotypes LG1:28146922
2122 (TT), LG1:28146925 (CC) and LG1:28146966 (GG) weigh on average 315.7g.
2123 ANNOVA tests between these three SNP markers and bodyweight at 220 dph

2124 showed a significant difference between fast growing and slow growing fish at
2125 220 dph at $P < 0.05$. Association data between bodyweight and SNPs are
2126 summarized in Supplementary Table 1.

2127

2128 **4.3 Brain Transcriptome**

2129 **4.3.1 Sexing tilapia with microsatellite genotyping**

2130 To differentiate the sex of sexually immature tilapia fries, I genotyped the
2131 parental pair and 42 sexually matured individuals at 120 days post hatch from the
2132 same cohort of tilapia raised for this RNA-seq experiment. All 42 sexually
2133 matured individuals had distinct external sex organ phenotype, with female
2134 oviduct and male genital papilla clearly distinguishable from each other. Using
2135 four microsatellite markers flanking the sex determining region, I reconstructed
2136 the haplotype of the parents to identify the alleles that are originating from the
2137 Y-chromosome. I found no recombination in the region within genotypes at these
2138 four microsatellite loci, suggesting that they are tightly linked loci flanking the
2139 sex-determining locus. The parental haplotypes reconstructed are shown in
2140 Figure 16.

2141
2142 Figure 16. Haplotypes of sex determining region for Mozambique tilapia at
2143 chromosome LG1. Numbers represent the size (bp) of PCR products flanking the
2144 microsatellite markers. The sex determining locus is located between S1543 and
2145 S001 on the Y-chromosome.

2146

2147 Using the information from the haplotype reconstruction, I genotyped the
2148 collected fin clips of all individuals used for RNA-seq at 5 dph, 30 dph and 90
2149 dph. The results of microsatellite genotyping can be found at Supplementary
2150 Table 2.

2151 **4.3.2 RNA-seq read statistics and mapping efficiency.**

2152 I generated 24 libraries from samples collected at three different development
2153 time points, namely 5 dph, 30 dph and 90 dph. These three time points are crucial
2154 time points during sexual development. At 5 dph, the sex determining
2155 mechanisms in tilapia were initiated²⁰⁹. At 30 dph, in female gonads, germ cell
2156 meiosis and oocyte formation were initiated^{37,209,210}. At 90 dph, in the male
2157 gonads, germ cell meiosis and spermatogenesis were initiated^{37,209,210}. In total,
2158 746 million 2 x 150 read pairs comprising of 223.8 gigabases of clean sequencing
2159 results were obtained after quality control with Trimmomatic¹⁸¹. Reads mapped
2160 to the reference genome O_niloticus_UMD_NMBU achieved an average
2161 mapping efficiency of 66% resulting in 20.8 million mapped read pairs. The raw
2162 fastq files were uploaded to DDBJ repositories with the BioSample Accession
2163 SAMD00195084-SAMD00195107.

2164

2165 Table 10. Summary of RNA-seq statistics for the brain transcriptome at 5 dph,
 2166 30 dph and 90 dph. The number of read pairs, mapping efficiency and the number
 2167 of mapped read pairs per library is indicated.

Developmental Group	Index	Number of reads (millions)	Mapping Efficiency (%)	Number of mapped read pairs (millions)
5 dph	5_F1	27.9	69.5	19.4
	5_F2	25.5	64.4	16.4
	5_F3	28.0	72.5	20.3
	5_F4	56.0	65.4	36.6
	5_M1	32.1	63.7	20.4
	5_M2	39.6	66.9	26.5
	5_M3	25.3	67.8	17.2
	5_M4	26.3	71.6	18.8
30 dph	30_F1	38.3	70.3	26.9
	30_F2	30.1	62.2	18.7
	30_F3	36.8	69.5	25.6
	30_F4	29.3	60.6	17.8
	30_M1	36.4	51.3	18.7
	30_M2	27.6	65.8	18.2
	30_M3	38.2	74.4	28.4
	30_M4	25.0	69.9	17.5
90 dph	90_F1	13.8	68.3	9.4
	90_F2	37.8	63.6	24.0
	90_F3	30.5	59.6	18.2
	90_F4	27.5	61.9	17.0
	90_M1	28.1	67.6	19.0
	90_M2	29.0	68.3	19.8
	90_M3	18.2	66.6	12.1
	90_M4	38.7	58.6	22.7
Average		31.6	66.0	20.8

2168

2169

2170 4.3.3 qPCR validation of RNA-seq data from brain tissues.

2171 I selected 12 genes including 10 protein coding genes and 2 non-coding RNAs
2172 for qPCR validations to validate the sex-biased genes identified in our
2173 transcriptomic analysis. Gene expression data were normalized using actin as
2174 housekeeping gene control. Log₂ fold-change result for each gene obtained from
2175 RNA-seq quantification and qPCR quantification were compared. An R-squared
2176 value of 0.9304 were obtained for the brain transcriptome comparison. The linear
2177 correlation curve was significantly non-zero at $p < 0.001$. The correlation linear
2178 curve can be found in Supplementary Figure 2.

2179

2180 4.3.4 Identification of sex-biased gene expression in tilapia brain

2181 Sex-biased gene expression analysis for all three developmental stages was
2182 conducted separately as three different DEseq2 comparisons¹³². The threshold for
2183 genes to be considered sex-biased was Benjamini & Hochberg corrected p -value
2184 ($padj$) < 0.05 and a fold-change absolute value of more than 2. In the 5 dph
2185 comparison group, 124 sex-biased genes were detected. 105 genes were
2186 upregulated in males and 18 were upregulated in females. In the 30 dph
2187 comparison group, 55 sex-biased genes were detected. Forty-four genes were
2188 upregulated in males and 11 were upregulated in females. In the 90 dph
2189 comparison group, 2706 sex-biased genes were detected, 2489 genes were
2190 upregulated in males while 217 genes were upregulated in females. The list of
2191 sex-biased genes can be found at [https://github.com/kentawan/omo-brain-](https://github.com/kentawan/omo-brain-RNAseq)
2192 [RNAseq](https://github.com/kentawan/omo-brain-RNAseq).

2193 In the 5 dph comparison group, 117 sex-biased genes are protein coding
2194 genes, 3 genes are long non-coding RNAs: two ribosomal RNAs and one small
2195 nuclear RNA. In the 30 dph comparison group, there are 51 protein coding genes,
2196 three long non-coding RNAs and one ribosomal RNA. In the 90 dph comparison
2197 group, there are 2004 protein coding genes, 627 long non-coding RNAs, 40
2198 ribosomal RNA and 32 transfer RNA.

2199 In this thesis, I am interested in finding neuroendocrine factors that may
2200 be sex-biased in expression and can regulate energy homeostasis and growth in
2201 tilapia. From our sex-biased genes, one candidate gene that may be responsible
2202 for sexual size dimorphism in tilapia is the pro-opiomelanocortin (*POMC*) gene.
2203 At 5 dph, *POMC* expression level did not show any difference between male and
2204 female tilapia. At 30 dph and 90 dph, expression of *POMC* was consistently
2205 female-biased with 8-fold and 32-fold expression level, respectively, in
2206 comparison to male brain tissue. The sex-biased expression of *POMC* at these
2207 two stages coincides with the beginning of sexual size dimorphism in
2208 Mozambique tilapia. Hence, I selected this gene for downstream gene function
2209 characterization in zebrafish. I selected the zebrafish as a model for *POMC*
2210 knockout functional analysis so as to take advantage of the shorter mutant
2211 generation time in comparison with tilapia. Another reason was that *POMC* has
2212 been shown to regulate feeding behavior in mice and human. It is easier to
2213 conduct feeding behavior assays in zebrafish compared to tilapia as zebrafish are
2214 smaller in size and various feeding behavior assays are well established for
2215 zebrafish.

2216

2217 Figure 17. MA plot of 5 dph sex-biased genes in the tilapia brain. Red dots
 2218 represent female-biased genes while blue dots represent male-biased genes.
 2219 Labelled genes represent the genes with the highest fold-change. Sex-biased
 2220 genes are defined as genes with $\log_2(\text{fold-change})$ of more than 1 and FDR-
 2221 adjusted $P < 0.05$.

2222

2223

2224 Figure 18. MA plot of 30 dph sex-biased genes in the tilapia brain. Red dots
 2225 represent female-biased genes while blue dots represent male-biased genes.
 2226 Labelled genes represent the genes with the highest fold-change. Sex-biased
 2227 genes are defined as genes with $\log_2(\text{fold-change})$ of more than 1 and FDR-
 2228 adjusted $P < 0.05$. The candidate gene *POMC* are labelled on the plot.

2229

2230

2231 Figure 19. MA plot of 90 dph sex-biased genes in the tilapia brain. Blue dots
 2232 represent male-biased genes while red dots represent female-biased genes.
 2233 Labelled genes represent the genes with the highest fold-change. Sex-biased
 2234 genes are defined as genes with $\log_2(\text{fold-change})$ of more than 1 and FDR-
 2235 adjusted $P < 0.05$. The candidate gene *POMC* are labelled on the plot.

2236

2237 4.3.5 Chromosomal locations of sex-biased genes.

2238 One hallmark of sex-biased genes is that they have non-random distribution
 2239 across all the chromosomes. I extracted the chromosomal location of each sex-
 2240 biased gene to test if the sex-biased genes identified in this study follow the non-
 2241 random distribution on chromosome observed in other species. Using the number
 2242 of observed sex-biased genes per chromosome at three different developmental
 2243 stages, I performed a Chi-Square test to test if the genes are randomly or non-
 2244 randomly distributed across the different chromosomes^{97,112,115}. The expected
 2245 number of genes per chromosome calculation stated above were derived from the
 2246 assumption that if a gene is randomly distributed across the genome, the number

2247 of sex-biased genes should be proportional to the number of genes on each
2248 chromosome. Chi-square results showed $p < 0.01$ for 5 dph sex-biased genes and
2249 $p < 0.001$ for 30 dph and 90 dph sex-biased genes, supporting the observations
2250 that sex-biased genes identified in this study are not randomly distributed across
2251 the genome.

2252

2253

2254 Figure 20. Chromosomal locations of observed and expected number of sex-
2255 biased genes in 5 dph tilapia.

2256

2257

2258 Figure 21. Chromosomal locations of observed and expected number of sex-
 2259 biased genes in 30 dph tilapia.

2260

2261

2262 Figure 22. Chromosomal locations of observed and expected number of sex-
 2263 biased genes in 90 dph tilapia.

2264

2265 4.3.6 Gene ontology and KEGG pathways of sex-biased genes.

2266 To understand the gene functions of the sex-biased genes identified in this study,
 2267 I used discontinuous MegaBLAST to assign zebrafish orthologues to every gene
 2268 in the tilapia genome with a cut-off threshold of E-value less than 0.001. Using
 2269 the matched orthologues, I performed gene ontology enrichment analysis for
 2270 biological processes using clusterProfiler in R¹⁸³. Using all *D. rerio* genes as
 2271 background and setting a Bonferroni corrected p-value of 0.05 as significant
 2272 threshold, I identified several significantly overrepresented gene ontology groups
 2273 for sex-biased genes identified in each different time point. For the 5 dph group,

2274 major gene ontologies identified include sensory system development, head
2275 development, brain development, sensory organ morphogenesis and epidermis
2276 development. For 30 dph comparison group, sex-biased genes were enriched in
2277 ontology groups including sensory system development, morphogenesis of an
2278 epithelium, connective tissue development and ossification. For 90 dph
2279 comparison group, sex-biased genes are enriched in gene ontology groups
2280 including nervous system process, sensory perception, sensory organ
2281 morphogenesis, ear development and cell fate commitment. A graphical
2282 summary of gene ontologies for brain sex-biased genes can be found at Figure
2283 23-25.
2284

2284

2285

2286 Figure 23. ClusterProfiler diagram summarizing the overrepresentations of gene
2287 ontologies in sex-biased genes identified in 5 dph tilapia brain tissue.

2288

2289

2290

2291 Figure 24. ClusterProfiler diagram summarizing the overrepresentations of gene
 2292 ontologies in sex-biased genes identified in 30 dph tilapia brain tissue.

2293

2294

2295

2296 Figure 25. ClusterProfiler diagram summarizing the overrepresentations of gene
2297 ontologies in sex-biased genes identified in 90 dph tilapia brain tissue.

2298

2299 Other than gene ontologies, I also investigated the pathways involved in
2300 sexual size dimorphism. Using the orthologues identified in zebrafish, I uploaded
2301 the sex-biased genes for each developmental stage onto PantherDB to test for
2302 overrepresentation in various pathways curated on the Panther Pathway database.
2303 In the 5 dph comparison group, pathways overrepresented by sex-biased genes
2304 include histamine H2 receptor mediated signaling pathway, presenilin pathway,
2305 thyrotrophin-releasing hormone receptor signaling pathway, cadherin signaling
2306 pathway, nicotinic acetylcholine receptor signaling pathway and Wnt signaling
2307 pathway. At 30 dph group, pathways overrepresented by sex-biased genes
2308 include Integrin signaling pathway, gonadotrophin-releasing hormone receptor
2309 pathway, presenilin pathway, nicotinic acetylcholine receptor signaling
2310 pathway, chemokine and cytokine signaling pathway and also pathways related
2311 to Huntington disease. In particular, at 30 dph, as the Mozambique juveniles are

2312 already sexually determined, but not yet sexually mature, I observed that genes
2313 from the gonadotrophin releasing hormone (GnRH) pathway are overrepresented,
2314 suggesting that at 30 dph there is already significant sexual differentiation in the
2315 brain.

2316 Table 11. Summary of Panther Pathways enrichment test for sex-biased genes in 5 dph tilapia brain transcriptome

PANTHER Pathways	No. of Background	No. of Hits	Expected Hits	Fold Enrichment	P value ²³¹⁷
Histamine H2 receptor mediated signalling pathway	34	2	0.31	6.47	4.17E-02 ²³¹⁸
Alzheimer disease-presenilin pathway	145	6	1.32	4.55	2.53E-03 ²³¹⁹
Thyrotropin-releasing hormone receptor signalling pathway	78	3	0.71	4.23	3.69E-02 ²³²⁰
Cadherin signalling pathway	166	6	1.51	3.97	4.79E-03 ²³²¹
Nicotinic acetylcholine receptor signalling pathway	123	4	1.12	3.58	2.81E-03 ²³²²
Wnt signalling pathway	348	8	3.17	2.53	1.59E-02 ²³²³

2325 Table 12. Summary of Panther Pathways enrichment test for sex-biased genes in 30 dph tilapia brain transcriptome

PANTHER Pathways	No. of Background	No. of Hits	Expected Hits	Fold Enrichment	P value
Integrin signalling pathway	232	24	7.09	3.38	7.69E-07
Gonadotropin-releasing hormone receptor pathway	301	23	9.2	2.5	1.69E-04
Alzheimer disease-presenilin pathway	145	11	4.43	2.48	6.97E-03
Nicotinic acetylcholine receptor signalling pathway	123	9	3.76	2.39	1.70E-02
Inflammation mediated by chemokine and cytokine signalling pathway	319	20	9.75	2.05	3.36E-03
Huntington disease	177	11	5.41	2.03	2.84E-02

2326

2327

2328

2329

2330

2331 Table 13. Summary of Panther Pathways enrichment test for sex-biased genes in 90 dph tilapia brain transcriptome

PANTHER Pathways	No. of Background	No. of Hits	Expected Hits	Fold Enrichment	P value
Circadian clock system	18	7	1.88	3.72	6.99E-03
Plasminogen activating cascade	13	5	1.36	3.68	2.26E-02
Heterotrimeric G-protein signalling pathway-rod outer segment phototransduction	50	15	5.22	2.87	8.98E-04
Histamine H2 receptor mediated signalling pathway	34	9	3.55	2.53	1.75E-02
Corticotropin releasing factor receptor signalling pathway	47	12	4.91	2.44	1.12E-02
5HT4 type receptor mediated signalling pathway	44	11	4.6	2.39	1.70E-02
Histamine H1 receptor mediated signalling pathway	53	13	5.53	2.35	9.54E-03
Metabotropic glutamate receptor group II pathway	62	15	6.47	2.32	5.67E-03
Blood coagulation	46	11	4.8	2.29	2.00E-02
Beta1 adrenergic receptor signalling pathway	59	14	6.16	2.27	8.55E-03
Muscarinic acetylcholine receptor 1 and 3 signalling pathway	81	19	8.46	2.25	2.97E-03
Nicotinic acetylcholine receptor signalling pathway	123	28	12.85	2.18	6.38E-04
Beta3 adrenergic receptor signalling pathway	40	9	4.18	2.15	4.59E-02
Beta2 adrenergic receptor signalling pathway	60	13	6.27	2.07	2.46E-02

Opioid proopiomelanocortin pathway	56	12	5.85	2.05	3.41E-02
Oxytocin receptor mediated signalling pathway	73	15	7.62	1.97	2.58E-02
Thyrotropin-releasing hormone receptor signalling pathway	78	16	8.15	1.96	1.98E-02
Muscarinic acetylcholine receptor 2 and 4 signalling pathway	81	16	8.46	1.89	2.35E-02
Cytoskeletal regulation by Rho GTPase	102	20	10.65	1.88	1.28E-02
Metabotropic glutamate receptor group III pathway	98	19	10.23	1.86	1.70E-02
5HT2 type receptor mediated signalling pathway	83	16	8.67	1.85	3.61E-02
TGF-beta signalling pathway	132	23	13.79	1.67	2.72E-02
Gonadotropin-releasing hormone receptor pathway	301	51	31.43	1.62	2.33E-03
CCKR signalling map	209	33	21.83	1.51	3.50E-02
Wnt signalling pathway	348	54	36.34	1.49	9.67E-03
Integrin signalling pathway	232	35	24.23	1.44	4.55E-02
Inflammation mediated by chemokine and cytokine signalling pathway	319	46	33.31	1.38	4.69E-02

2332

2333 At 90 dph, I observed a large number of pathways (27 pathways) enriched
2334 with sex-biased genes in comparison with 30 dph (6 pathways) and 5 dph (6
2335 pathways). These pathways include pathways related to neuronal functions and
2336 synapses, hormonal pathways and G-protein signaling pathways. Sex-biased
2337 neuronal pathways include nicotinic acetylcholine receptor signaling pathway,
2338 phototransduction pathway, muscarinic acetylcholine receptor pathway and
2339 5HT4 type receptor mediated signaling pathway. The majority of these pathways
2340 are involved in neuronal synapse signal transduction. Besides neuron-related
2341 pathways, pathways related to G-protein signaling pathways were found to be
2342 sex biased in expression as well. These include beta adrenergic receptor signaling
2343 pathways, metabotropic glutamate receptor pathway, histamine H1 receptor
2344 mediated pathway and chemokine and cytokine signaling pathways. These
2345 pathways are involved in the transduction of extracellular signals to initiate
2346 downstream effectors intracellularly, such as transcription, motility, contractility,
2347 and secretion²¹¹.

2348 I also identified four sex-biased neuroendocrine pathways, especially
2349 those that are localized in the pituitary glands. These pathways include the GnRH
2350 pathway, corticotrophin releasing factor receptor signaling pathway,
2351 thyrotrophin-releasing hormone receptor signaling pathway and the opioid
2352 proopiomelanocortin pathway. These pathways are involved in the
2353 hypothalamus-pituitary-gonadal (HPG) axis, hypothalamus-pituitary-adrenal
2354 (HPA) axis and the hypothalamus-pituitary-thyroid (HPT) axis²¹². These axes
2355 relay molecular cues from the central nervous system into the pituitary glands,
2356 which secrete neuroendocrine hormones such as growth hormones, thyroid

2357 stimulating hormones, gonadotrophins and proopiomelanocortin hormones
2358 (*POMC*)^{169,212,213}.

2359 **4.3.7 Non-synonymous/Synonymous rate of mutations of sex-** 2360 **biased genes**

2361 Sex-biased genes are hypothesized to be the product of resolving sexual conflict
2362 during evolution^{4,93,96,97,107,108}. In various species, sex-biased genes evolve
2363 rapidly in the protein sequences, hence accumulating higher numbers of non-
2364 synonymous relative to synonymous mutation sites, i.e. higher *dN/dS* ratio
2365 ^{4,96,97,108}. I evaluated the *dN/dS* ratio of sex-biased genes in the brain by
2366 comparing the coding sequences of tilapia with closely related cichlid species
2367 including *Maylandia zebra*, *Amphilophus citrinellus*, *Neolamprologus brichardi*,
2368 *Astatotilapia burtoni* and *Haplochromis nyererei*. I also added two species as
2369 outgroup to increase our detection power of positive selection. These two species
2370 are *Tetraodon nigroviridis* and *Oryzias latipes*. I selected 235 genes that are non
2371 sex-biased ($p > 0.90$, $\log_2$ fold-change < 1) in all three developmental stages (5
2372 dph, 30 dph and 90 dph) as unbiased gene control group. Using protein alignment
2373 guided codon alignment from PAL2NAL, I investigated the *dN/dS* ratio of sex-
2374 biased genes at all three developmental stages and also unbiased genes using the
2375 PAML software suite¹³⁶. The median *dN/dS* for unbiased genes are 0.176,
2376 suggesting a strong purifying selection on these genes and deleterious mutations
2377 are removed during evolution. The median *dN/dS* of female biased genes are
2378 0.253, 0.254 and 0.275 for 5 dph, 30 dph and 90 dph respectively. The median
2379 *dN/dS* of male biased genes are 0.228, 0.329 and 0.244 for 5 dph, 30 dph and 90

2380 dph, respectively. dN/dS of male-biased genes and female-biased genes are
2381 significantly higher than unbiased genes (Mann-Whitney U Test; $P < 0.0001$) for
2382 5 dph, 30 dph and 90 dph. This result suggests that some of the sex-biased genes
2383 are under relaxed purifying selection, which allows the accumulation of non-
2384 synonymous mutations in the codons of sex-biased genes, leading to protein
2385 sequence changes and possibly functional divergence.

2386 To test if the higher dN/dS values of sex-biased genes is caused by an
2387 increase of dN value but not a decrease in dS values, I extracted the dN and dS
2388 values for each comparison group and performed a Mann-Whitney U Test. I
2389 found that all sex-biased groups have significantly higher dN values ($> 14\%$
2390 differences) compared to the unbiased genes (Mann-Whitney U Test; $P < 0.0001$).
2391 However, I found only a small ($< 7\%$) difference in dS values of 30 dph males
2392 and 90 dph females (Mann-Whitney U Test; $P < 0.001$) while the remaining four
2393 groups showed no significant differences in dS values. Hence, it may be inferred
2394 that the higher dN/dS values in sex-biased genes are caused by the increase of
2395 non-synonymous mutations instead of a decrease of synonymous mutations.

dN/dS of Sex-biased Genes

2396

2397 Figure 26. The dN/dS ratio of sex-biased genes identified from tilapia brain transcriptomes

2398

2399

2400 Figure 27. dN/dS ratio of sex-biased genes in various developmental stages categorized into 0.1 intervals. More than 50% of unbiased genes have
2401 dN/dS ratio of less than 0.2 while sex-biased genes on all developmental stages have higher dN/dS on average.

2402 **4.3.8 Identification of SNPs from RNA-seq data**

2403 Single nucleotide polymorphisms (SNPs) are one of the most abundant classes
2404 of polymorphisms in animal genomes. In comparison to markers such as simple
2405 sequence repeats (SSRs), SNPs are usually bi-allelic, offer lower error rates in
2406 allele calling and occur at shorter marker intervals²¹⁴. SNPs located in the exonic
2407 region may cause non-synonymous mutations and lead to amino acid changes or
2408 premature stop codons. To call SNP alleles from our transcriptomic data, I
2409 aligned RNA-seq reads to the reference genome using a 2-pass alignment method
2410 in STAR. After preprocessing to remove duplicates, rearrangement of indels and
2411 recalibration of base scores, all variants including SNPs and indels were called
2412 with samtools using a cut-off alignment score Q30 and minimal coverage depth
2413 of 20x coverage. In total, 2,701,123 SNPs and 311,826 indels were genotyped
2414 from 12 males and 12 females' transcriptomic data. 2,940,483 SNPs and indels
2415 were located in assembled chromosomes and 66,627 were located in scaffold
2416 positions. From the total number of SNPs called, 24,149 SNP sites were found to
2417 be multi-allelic.

2418

2419 **4.3.9 Association of SNP markers in chromosome LG1 with sex**

2420 Using the markers called from our samples, I performed an association study
2421 between the phenotype of sex and markers called from each sample. I omitted
2422 eight samples with significant amount of missing data and proceeded with
2423 association study using 16 samples in PLINK. The SNP association results are
2424 summarized in Figure 28. Based on the result output of PLINK, the peak SNP is

2425 significant at chromosome wide threshold ($P\text{-val} = 1.18\text{E-}05$). This SNP is
2426 located at chrLG1: 22,948,704, which is located in the intron region of an
2427 uncharacterized non-coding RNA LOC106098012. Two other markers flanking
2428 this SNP are found at chrLG1: 22,547,726 ($P = 0.00276$) and chrLG1: 23,009,787
2429 ($P = 0.00276$). This region is 462 kbp in length and contains 28 genes, including
2430 26 protein coding genes and 2 non-coding RNA.

2431

2432

2433 Figure 28. Association of SNPs with the phenotype of sex in chromosome LG1. Red line represents genome-wide significance threshold while blue line
 2434 represents chromosome-wide significant threshold.

2435 4.4 *In vitro* genome editing of candidate genes in OmB cell lines

2436

2437 I attempted to generate knock-out isoclonal OmB cell lines for three candidate
2438 genes using pSpCas9(BB)-2A-GFP plasmid, namely *LPINI*, *RASGRF1* and
2439 *POMC*. *LPINI* and *RASGRF1* were female-biased and male-biased genes,
2440 respectively, identified from the skeletal muscle transcriptome dataset while
2441 *POMC* was identified from the brain transcriptome dataset. *LPINI* is responsible
2442 for the final step of triglyceride synthesis and may be responsible for cell growth
2443 through regulating fatty acid metabolism and adipocyte differentiation²¹⁵.
2444 *RASGRF1*, on the other hand, has been shown to delay growth in mice when
2445 knocked-out *in vivo*²¹⁶. *POMC* mutations in humans, mice and dogs have been
2446 reported to be associated with obesity and increased linear growth^{168,169,172,213,217}.
2447 As such, these three sex-biased genes identified from our RNA-seq datasets were
2448 selected as candidate genes for *in vitro* genome editing. Cells with weak GFP
2449 expression were observed at 48 hours post transfection for all constructs targeting
2450 *RASGRF1*, *LPINI* and *POMC*. Using wide-field microscopy on an inverted
2451 microscope, transient GFP signals can be observed from 48 hours post
2452 transfection up to one-week post transfection. However, the transfection
2453 efficiency was very low at 6.15% on average for all targets. I repeated the
2454 transfection with less plasmid DNA at 500ng and 1000ng but the transfection
2455 efficiency remains low at 3.31% and 3.78% respectively. Using T7 endonuclease
2456 assay, I was not able to detect a double band on the PCR gel.

2457 To troubleshoot for the possibility of poor transfection due to wrong
2458 transfection protocol, I attempted to transfect OmB cells with pEGFP-N1-FLAG
2459 plasmid using the same protocol for pSpCas9(BB)-2A-GFP plasmid transfection.

2460 The plasmid pEGFP-N1-FLAG is a smaller plasmid at 4.4kb with a
2461 cytomegalovirus promoter (pCMV) driving the expression of an eGFP protein.
2462 Using the same protocol as mentioned for pSpCas9(BB)-2A-GFP plasmid
2463 transfection, transfected OmB cells were able to achieve ~32.6% transfection
2464 efficiency, suggesting that OmB cells can be transfected successfully using
2465 Lipofectamine reagents. These results suggest that the pSpCas9(BB)-2A-GFP
2466 plasmid, which was originally constructed for mammalian cell lines such as
2467 HEK293FT cells and HUES9 cells, were not suitable for genome-editing in
2468 teleost cell lines¹⁴⁷.

2469

2470 **4.5 *In vivo* genome editing of *pomc* gene in zebrafish**

2471 **4.5.1 *POMC* in Mozambique tilapia**

2472 I choose *POMC* as a promising candidate gene for understanding the mechanisms
2473 of sexual size dimorphism in tilapia because of the following reasons. Firstly,
2474 *POMC* is a central energy homeostasis gene involved in regulating energy
2475 homeostasis and bodyweight control, as has been demonstrated in mice^{172,218}. As
2476 sexual size dimorphism is inherently linked with differential growth rate and
2477 energy homeostasis, *POMC* is a promising candidate gene for achieving sexual
2478 size dimorphism in tilapia.

2479 Secondly, *POMC* was found to be consistently female-biased in
2480 expression at 30 dph and 90 dph female tilapia, which coincides with the stage
2481 where sexual size dimorphism is developing in tilapia. In our qPCR data, I
2482 verified that *POMC* is consistently female-biased in expression from 30 dph to

2483 120 dph. Thirdly, *POMC* has been shown to regulate feeding behavior in mice
2484 and dog¹⁷⁰. Tilapia being a maternal mouth brooding species, may have female-
2485 biased *POMC* expression in the brain to suppress the feeding behavior of females
2486 during periods of mouth brooding behavior.

2487 The *POMC* gene is located on chromosome LG23 in the tilapia genome.
2488 It has three exons with a total mRNA length of 908 bp, a coding sequence length
2489 of 627 bp and a translated protein length of 208 amino acids. *POMC* is a
2490 polypeptide hormone. After translation, it is further cleaved into different
2491 peptides, namely *ACTH*, alpha melanocyte stimulating hormone, beta
2492 melanocyte stimulating hormone and beta endorphin. Brain tissue qPCR on 30
2493 dph, 90 dph and 120 dph sexually matured male and female Mozambique tilapia
2494 showed that *POMC* is highly expressed at whole brain tissue and is female-biased
2495 in expression ($P < 0.05$) (Figure 29).

2496

2498

2499 Figure 29. Expression level of *POMC* in the brain from 5 dph to 120 dph as
2500 validated by qPCR from brain tissues of tilapia (N = 4 for each sex at each time-
2501 point).

2502

2503

2504 **4.5.2 Possible functional estrogen receptor elements upstream**
2505 **of *POMC* transcriptional start site**

2506 Using JASPAR, which is a database of curated and non-redundant transcription
2507 factor and enhancer binding motifs, I searched for binding site motifs associated
2508 with estrogen receptors or estrogen related receptors. I identified two putative
2509 EREs located at 2388 bp and 1379 bp upstream of *POMC* transcriptional start
2510 site in tilapia genome, at 86% and 88% similarity to mammalian EREs,
2511 respectively. The sequence of these two predicted EREs are 5'-
2512 AGGTCACAGTAGGCA-3' and 5'-TGTACACCCTGACCC-3', respectively.
2513 These two predicted EREs were amplified as one 1.5 kb PCR fragment and

2514 cloned into a pGL3 Promoter vector, which contains a SV40 promoter upstream
2515 of a luciferase coding sequence for testing putative enhancer region. OmB cells
2516 were transfected with *pomc*ERE-pGL3-promoter using Lipofectamine and
2517 subsequently challenged with β -Estradiol. In the luciferase assay, I found that β -
2518 Estradiol can significantly enhance the expression of Luciferase activity using
2519 this induction test. In transfected OmB cells treated with ethanol as control, the
2520 median Renilla-normalized Luciferase activity was 1999 relative luminometer
2521 units (RLU) at 48 hours post transfection. In transfected OmB cells treated with
2522 a gradient of 1 nM, 10 nM and 100 nM β -Estradiol, the median Renilla-
2523 normalized Luciferase activity was 3447 RLU, 4019 RLU and 2617 RLU at 48
2524 hours post transfection, respectively. This corresponds to 1.72, 2.01 and 1.31
2525 fold-change induction respectively for 1 nM, 10 nM and 100 nM β -Estradiol in
2526 comparison to ethanol control. The higher induction of luciferase activity by β -
2527 Estradiol in OmB cells was significant at $P < 0.05$ for all three concentrations of
2528 β -Estradiol. Luciferase activity in ethanol control and estradiol treatment is
2529 summarised in Figure 30.

2530

2531 Figure 30. Induction of Luciferase activity in *pomc*ERE-pGL3-promoter
 2532 transfected OmB cells in ethanol treatment and β-Estradiol treatment (1 nM, 10
 2533 nM and 100 nM) at 48 hours post transfection. Luciferase activity shown are
 2534 normalised with Renilla expression control. (* $P < 0.05$)

2535 4.5.3 Establishment of a *POMC*-knockout zebrafish line as a 2536 model to study somatic growth and feeding behavior

2537 To study the role of the candidate gene *POMC* *in vivo* with relation to growth, I
 2538 generated a zebrafish *POMC* knockout line. I designed three sg-RNAs targeting
 2539 three different exonic regions of *POMC* in zebrafish, namely *pomc-t1*, *pomc-t2*
 2540 and *pomc-t3*. All three sgRNAs were co-injected with Cas9 protein into 480
 2541 single cell stage embryos in a multiplex CRISPR/Cas9 reaction setup. 268
 2542 embryos survived to the larval stage, while the remaining embryos were either
 2543 not fertilized or did not develop successfully. Embryos were transferred to a
 2544 circulatory tank system at 15 dph and raised to sexually matured stage. I observed

2545 18 embryos with significant defects, especially missing the posterior part of the
2546 body trunk. These embryos did not survive to adulthood. T7 endonuclease assay
2547 showed that these embryos did not have successful genome editing event at target
2548 sites, suggesting that they could be a phenotype of off-target effect or a side-
2549 effect from microinjections. Only 58 zebrafish survived to sexually matured stage
2550 at 90 days post hatch. Fin clips were incised for genotyping after the fish were
2551 anaesthetized with MS-222 and body trunk photos were taken using Leica
2552 stereomicroscope.

2553 To genotype the injected fish, I amplified the 650 bp region flanking each
2554 target genome editing sites. PCR fragments were used for T7 Endonuclease assay.
2555 Individuals with positive genome editing events will show two cleaved bands
2556 between 310 bp and 330 bp region. Through direct Sanger sequencing of the 650
2557 bp PCR product, I obtained “double peaks” following the known gRNA target
2558 sites, indicating that indels had occurred and interrupted the ability of capillary
2559 Sanger sequencing to resolve the individual bases.

2560

2561

2562

2563 Figure 31. T7 genotyping assay of F0 fishes. Fish number 4 showed no indels at
2564 target site *pomc*.

2565

2566
2567

2568

2569 Figure 32. Direct sequencing of PCR product amplified from 600 bp region
2570 flanking the target site. Lane 1 and 2 show successful genome editing with
2571 characteristic “double peaks” while lane 3 and 4 showed wild-type genotype.

2572

2573 Out of 58 F0 fish, 18 fish showed successful cleavage with T7
2574 endonuclease I assay. To identify the allele frequency of successful F0 knockout
2575 mutants, I pooled the PCR products of 13 fish and cloned them into T-vector.
2576 After transformation, 192 white colonies were picked for colony PCR and
2577 sequenced. I found a mosaic genotype of 9 different indels as a product of non-
2578 homologous end joining events following Cas9 protein cleavage. 52.6% of all
2579 alleles genotyped contained indels while the remaining were still wild types.

2580

Genotype																																					Allele Frequency (%)													
1.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	-	-	C	C	G	G	C	T	A	G	C	A	A	G	C	G	C	T	A	T	G	G	A	G	G	C	T	6.3		
2.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	C	C	G	C	C	G	G	C	T	A	G	C	A	A	G	C	-	-	T	A	T	G	G	A	G	G	C	T	4.2	
3.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	C	C	G	C	C	G	G	C	T	A	G	C	A	A	-	-	-	-	-	-	G	G	A	G	G	C	T	4.7		
4.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	-	C	C	G	C	C	T	A	-	-	-	-	-	-	-	-	-	-	T	G	G	A	G	G	C	T	2.1			
5.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	C	C	G	C	C	G	G	C	T	A	G	C	A	A	G	C	-	-	-	-	-	-	G	G	A	G	G	C	T	2.6
6.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	G	C	T	T	A	C	C	G	C	C	G	G	C	T	A	G	C	A	A	G	C	-	-	-	-	-	-	G	G	A	G	G	C	T	7.8
7.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	-	C	C	G	-	-	-	-	-	-	-	-	-	A	G	C	G	C	T	A	T	G	G	A	G	G	C	T	16.7	
8.	C	A	T	T	C	C	G	C	T	G	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	-	G	G	C	T	A	G	C	A	A	G	C	G	C	T	A	T	G	G	A	G	G	C	T	8.3
WT.	C	A	T	T	C	C	G	C	T	G	G	A	G	C	G	T	C	-	-	-	C	C	G	C	C	G	G	C	T	A	G	C	A	A	G	C	G	C	T	A	T	G	G	A	G	G	C	T	47.4	

2581

2582 Figure 33. Estimated allele frequency of F0 mutants based on T-vector clones containing the 600 bp region flanking the genome editing target site
2583 *pomc*.

2584 Using the positive knockout mutants, I crossed the mutants to wildtype
2585 AB fish to test if I had successful germ line transmission of mutation. I collected
2586 15 dph whole embryos for DNA extraction and performed T7 Endonuclease
2587 assay. Out of the 32 embryos collected, eight embryos showed positive cleavage
2588 with T7 endonuclease assay. I also designed primers flanking the entire POMC
2589 gene in zebrafish to genotype the F1 mutants. I found two large deletion alleles
2590 of 381 bp and 1613 bp deletions. The 381 bp deletions correspond to the loss of
2591 fragments between the target sites of sgRNA2 and sgRNA3. The larger 1613 bp
2592 deletion corresponds to the loss of fragment between the target sites of sgRNA1
2593 and sgRNA2. These results confirm that our indel mutations achieved germline
2594 transmission.

2595

2596 Figure 34. PCR genotyping of F1 mutants showing large deletions of
2597 approximately 381 bp and 1613 bp (indicated by red boxes).

2598

2599 The expression level of *pomc* mutants were checked using quantitative
 2600 real-time PCR. Using GADPH as housekeeping gene normalization, I found that
 2601 the expression level of POMC in *pomc* mutants to be very low in comparison
 2602 with wild type zebrafish ($P < 0.01$). The GADPH normalized expression level of
 2603 POMC in mutant and wildtype were shown in Figure 35.

2604
 2605 Figure 35. Expression level of *POMC* in wildtype zebrafish and *pomc* mutants.
 2606 ($P < 0.01$)

2607 2608 **4.5.4 F0 mutants showed smaller melanophores distribution**

2609 I observed that fish with successful POMC genome editing events showed
 2610 smaller melanophore distribution in comparison to wild type fish. In some
 2611 instances, small patches without melanophores on the skin were observed. To
 2612 quantify the differences of melanophore distribution between *pomc* mutants and

2613 wildtype fish, I obtained lateral view images of zebrafish anterior trunk and
 2614 posterior trunk. Images were preprocessed and quantified for melanophore
 2615 distribution using R. On the anterior trunk, the average coverage area of
 2616 melanophores is 16.7% for the wildtype fish and 6.6% for F0 *pomc* mutants. On
 2617 the posterior trunk, the average coverage area of melanophores was 17.3% for
 2618 wildtype fish and 6.9% for F0 *pomc* mutants. For both anterior and posterior
 2619 trunk, *pomc* mutants showed significantly smaller melanophore distribution
 2620 compare to wild type fish (Two-tailed T-test; $P < 0.05$).

2621

2622

2623 Figure 36. Anterior lateral view of zebrafish F0 mutants showing phenotype of
 2624 small and less dense melanophores on F0 *pomc* mutants.

2625

2626

2627 Figure 37. Posterior lateral view of zebrafish F0 mutants showing phenotype of
 2628 small and less dense melanophores on F0 *pomc* mutants.

2629

2630

2631

2632 Figure 38. Quantitative analysis of melanophore distribution for the anterior and
2633 posterior trunk lateral view of F0 *pomc* zebrafish mutants and wildtype control.
2634 (* $P < 0.05$)

2635

2636 **4.5.5 Increased growth observed in F1 heterozygous *pomc***
2637 **zebrafish**

2638 I measured the body length and body weight of 10 mutant and 21 wildtype F1
2639 zebrafish at 80 dph. I then compared the total body length (mouth to tail) and
2640 body weight of F1 mutants and wildtype fish. For body length, mutant fish were
2641 7 mm longer than wild type fishes in median length ($P = 0.00158$, Kruskal-Wallis
2642 Test). For body weight, mutant fish were 138 mg heavier than wildtype fish in
2643 terms of median bodyweight ($P = 0.00934$, Kruskal-Wallis Test). The
2644 comparison results between mutant and wildtype fish are summarised in Figure
2645 39.

2646

2647

2648 Figure 39. Summary of phenotypic comparison between wildtype zebrafish and
 2649 *pomc* zebrafish for traits of body weight (A.) and body length (B.). The orange
 2650 histogram plot on the right shows the median difference at 95% confidence
 2651 interval. The median difference is plotted on a floating axis on the right as a
 2652 bootstrap sampling distribution. The median difference is depicted as a black dot;
 2653 the 95% confidence interval is indicated by the ends of the vertical error bar.

2654

2655 **4.5.6 *pomc* juveniles showed shorter feeding reaction to feeding in**
2656 **comparison to wildtype**

2657 In our phenotypic observations of *pomc* juveniles, I observed that *pomc* juveniles
2658 are more sensitive to feeding routine. *pomc* juveniles moved to the surface at a
2659 shorter time, after feed was provided, in comparison to wildtype fish. Using video
2660 capture analysis, I was able to quantify the feeding response latency differences
2661 between *pomc* and wildtype fishes. In my analyses, wildtype had a median
2662 feeding latency of 13.5 seconds while *pomc* fishes had a median feeding latency
2663 of 1.0 second. The median difference between *pomc* fish and wildtype fish was
2664 significant, at 12.5 seconds ($P < 0.001$, Kruskal-Wallis test). A summary of
2665 feeding latency differences is shown at Figure 40.

2666

2667

2668

2669 Figure 40. The median difference in feeding latency between wildtype and *pomc*
2670 zebrafish. Both groups are plotted on the left axes; the median difference is
2671 plotted on a floating axis on the right as a bootstrap sampling distribution. The
2672 orange histogram plot on the right shows the median difference at 99%
2673 confidence interval. The median difference is depicted as a black dot; the 99%
2674 confidence interval is indicated by the ends of the vertical error bar.

2675

2676 **5 Discussion and future works**

2677

2678 **5.1 Whole genome bisulfite sequencing**

2679 Epigenetics refers to the changes in gene expression levels without modifications
2680 of the genome sequences²⁰⁰. Epigenetics can be seen as an interesting explanation
2681 for sexual dimorphism, as individuals of two sexes from the same species share
2682 almost identical chromosomes except for the sex chromosomes^{61,202,219}. Through
2683 epigenetic mechanisms such as paternal or maternal imprinting, epigenetic
2684 silencing, chromatin inactivation or chromosomal inactivation mediated by non-
2685 coding RNA, species exhibiting sexual dimorphism may regulate gene
2686 expression levels in each sex, hence leading to phenotypic plasticity in different
2687 sexes^{102-104,121}.

2688 I showed through correlation analysis, the majority of CpG methylation
2689 profile in male and female tilapia were largely conserved. However, I was able
2690 to identify regions of sexually dimorphic DNA methylation levels within the
2691 tilapia genome. Correlation of arbitrary gene promoter regions with gene
2692 expression levels also revealed a negative correlation between promoter
2693 methylation and gene expression levels. I also showed that in genomic regions
2694 approaching transcriptional start sites, CpG methylation is low in tilapia genome.
2695 In other parts of the genome such as transposon regions and repeat regions, CpG
2696 methylation level is high. Similar observations have been reported in other
2697 vertebrate species including puffer fish, zebrafish and mouse^{201,202}. I also
2698 identified differentially methylated regions in chromosome LG1, which in
2699 previous publications, have been identified QTL associated with sex

determination and growth in tilapia. This phenomenon is also reported in stickleback, in which chromosome 19, the sex chromosome identified in stickleback, has higher number of hypermethylated CpG cytosine in comparison with other chromosomes²¹⁹.

I also inferred putative active cis-regulatory regions from our WGBS datasets, which may have important roles in regulating the gene expression of sex-biased genes in somatic tissues. Using WGBS data to infer and predict ACRs adds value to the WGBS datasets and allows more insights regarding the functional aspect of CpG islands with low methylation level. WGBS at single-base resolution in other models have revealed that regulatory regions with low counts of CpG sites generally acquire a low methylation state when occupied by transcription factors^{104,105,204,220-222}. Many of these cis-regulatory regions have unknown functions. However, the attribution of “function” to these low methylation state regions assumes a general repressive effect of gene expression level by CpG methylation within a particular regulatory region¹⁰⁵. SNPs or other genetic polymorphisms present in these cis-regulatory regions may be important in regulating the gene expression of genes corresponding to the cis-regulatory region.

2718

5.1.1 Advantage of single base resolution whole genome bisulfite sequencing over other methodology

In this study, I presented the first single base resolution CpG methylome in tilapia. Single base resolution whole genome bisulfite sequencing provides the most

comprehensive coverage in comparison with other methods such as reduced representation bisulfite sequencing (RRBS) or Methylated DNA Immunoprecipitation sequencing (MeDip-Seq). Using WGBS, I am able to cover up to 47.9% of all CpG sites at least 5x in coverage. In a systematic comparison study done using human stem cells, WGBS shows the highest coverage of CpG sites at 14.73% while RRBS and MeDip-Seq only covered 0.37% and 0.09% of all CpG sites in the human genome. Besides lower coverage of CpG sites, RRBS and MeDip-Seq also showed bias in coverage of CpG sites in the genome. RRBS relied on restriction enzymes which are sensitive to CpG methylation on target restriction sites while MeDip-Seq enriches for regions with lower CpG densities. However, the advantage of WGBS or MethylC-seq comes at a trade-off of up to 50 x increase in cost. This is due to the fact that MethylC-Seq is essentially a whole genome sequencing project with a quantitative hypothesis (CpG methylation %). Hence, given a finite budgetary constraint, I can only conduct two biological replicates for each sample, which fulfils the minimal requirement for computing statistical significance of differentially methylated CpG sites identified. Using other methods, I may be able to increase the number of biological replicates to increase the statistical power of DMRs identified.

5.1.2 Epigenetic architecture of tilapia is similar to other vertebrate animals.

Based on our mapping results, on average 69.6% of cytosines in CpG context are methylated. In CHH and CHG context, only 0.6% and 0.5% of the cytosines are methylated, respectively. This result is consistent with findings in other vertebrates (e.g. *Tetraodon nigroviridis*)²⁰², but is different from that in plants

(e.g. *Oryza sativa*)²⁰². CHH and CHG methylation is a major characteristic of plant methylomes and is largely absent or found in very low quantity in animal methylomes²⁰⁰⁻²⁰². In pufferfish, 65.5% of cytosines in CpG are methylated while cytosines in CHG and CHH are only methylated at 0.3% and 0.3%, respectively²⁰². Other animal species showed similar traits lacking methylation on CHH and CHG cytosines. In the rice methylome, only 50.00% of CpG is methylated while 27.40% and 5.20% of cytosines are methylated in CHG and CHH context^{201,202}.

The gene structure CpG methylation profile is consistent with other sequenced vertebrate methylomes^{202,223}, which highlights the rapid drops in CpG methylation level approaching the TSS, which is a region with many cis-regulatory elements as well as transcriptional binding site. Similar results were reported in plants^{202,223}, fish⁴⁰, and humans⁴³, suggesting that the CpG methylation profiles of vertebrates on the gene structure are largely conserved from fish to humans. Our results are consistent with the observations by Suzuki et al, who observed the global CpG methylation, except on CpG islands, as contrasted by the mosaic CpG methylation found in some plants (e.g. *Arabidopsis*), invertebrates (e.g. *C. intestinalis*) and fungi (e.g. *Neurospora*)¹⁰². High CpG methylation within the gene body may be involved in a proposed mechanism to prevent initiation on the gene body instead of TSS, so as to prevent the production of aberrant mRNAs²²⁴.

5.1.3 Repeat elements in tilapia genome are highly methylated.

In my WGBS analysis, I annotated repeat regions found in the tilapia genome and quantified the CpG methylation in these elements. I found that up to 16.51%

of the tilapia genome are made up of retroelements, repeat elements and DNA transposons. I found that more than half of the retroelements are CpG hypermethylated (> 75% methylation level) in both male and female tilapia. This result is consistent with observations in mammals and plants^{201,202}. Repeats and transposable elements may disrupt genome stability, and hence need to be silenced by DNA methyl transferases (DNMT) through CpG methylation²²⁵. In zebrafish, *DNMT1* mutants exhibit constant induction of RNA transcripts from class I retrotransposons due to hypomethylation of transposable elements and lack of CpG methylation silencing mechanisms on these transposable elements²²⁶. Thus, in the tilapia genome, high level of CpG methylation on these transposable elements may be interpreted as a means of genome-defense against transposable elements, hence ensuring genomic stability. The effect of CpG methylation on transcript levels of transposable elements in tilapia remains unknown, as our RNA-seq protocol only selectively investigates transcripts that are poly A-tailed.

2785

2786 **5.1.4 CpG methylation in promoter regions is correlated with gene** 2787 **expression level in tilapia**

I investigated the effect of CpG methylation in the promoter region of tilapia on the gene expression level. Based on our results, I found that CpG methylation levels in the promoter regions of tilapia are inversely correlated with gene expression level. Gene suppression effect is most robust when the number of CpG sites in the corresponding promoter region is high. This finding is consistent with reported vertebrate methylomes such as human peripheral blood mononuclear cells (PBMCs), mice and tetraodon fish^{102,201,202}. Our data extends the

2795 conventional view that promoter CpG methylation in teleost genomes can
2796 suppress gene expression, similar to the observations found in mammalian
2797 genomes²⁰¹. As such, the mechanism of suppressing gene expression via CpG
2798 methylation in the promoter regions may have an even earlier evolutionary origin
2799 than teleosts.

2800

2801 **5.1.5 Tilapia methylome is sexually dimorphic in somatic tissue**

2802 In this thesis, I defined the significance threshold of CpG methylation level
2803 difference as more than 25% difference. I selected 25% as our threshold because
2804 this methylation difference level gives better confidence on the biological
2805 significance of CpG methylation differences, although in some papers, only 10%
2806 methylation differences is used. Setting the threshold to the more conservative
2807 25% difference would also be more prudent in our case due to our relatively low
2808 minimum coverage (>5X per site). In my analysis, I found 1,717 CpG sites that
2809 were differentially methylated between male and female tilapia muscle tissue.
2810 This shows that although the tilapia epigenome is largely conserved between
2811 sexes, with a Pearson correlation of 0.86, there are still distinct CpG methylation
2812 sites that are differentially methylated. However, single base-pair methylation is
2813 often not informative enough to infer functions for DMRs on a genomic scale,
2814 partly due to missing data points which gave heavy penalties to the q-values when
2815 calculating differential methylation. Hence, I applied a sliding window analysis
2816 to average out the missing data points in CpG methylation sites.

2817 **5.1.6 Inference of sexually dimorphic active cis-Regulatory Elements**

2818 Encouraged by the development of robust algorithms to interpret high coverage
2819 WGBS data^{227,228}, I intended to explore the regulatory elements of hybrid tilapia
2820 with the aim of identifying non-coding regulatory elements that may provide
2821 evidence to explain the phenomenon of sexual size dimorphism. Using the
2822 WGBS data generated for tilapia, I inferred the active CREs from the hybrid
2823 tilapia genome using the software MethylSeekR. I defined the promoter as 500
2824 bp upstream of TSS in this section of our analyses as I was interested in
2825 transcription factor binding sites that are located close to the TATA-box, which
2826 is usually located 25-35 base pairs upstream of the transcription start site.
2827 Defining a promoter region that is too long (e.g. 1000 bp) in this analysis may
2828 lead to false positives. Previous studies have shown that transcription factors
2829 bound to cis-regulatory elements (CRE) found on the promoter regions upstream
2830 of TSSs will shape the CpG methylation levels on the CRE²²⁷. It is conventionally
2831 understood that highly methylated promoter regions are associated with
2832 suppressed gene expression and vice versa. Gene bodies, on the other hand, are
2833 often highly methylated when expression level is high²²⁹. In our results, I inferred
2834 sexually dimorphic active CREs in male and female tilapia muscle tissues. Using
2835 WGBS data to infer and predict active CREs add value to the WGBS datasets, as
2836 using WGBS to identify only CpG was often criticized as expensive and
2837 inefficient, as only CpG methylation base pairs from WGBS were used for
2838 analysis. With the predicted coordinates and sequence of CREs in tilapia, I may
2839 design experiments to functionally characterize them and understand their
2840 functions in sexual size dimorphism in tilapia.

2841

2842 **5.1.7 Future works in tilapia CpG methylation analysis**

2843 Although we can infer from our WGBS and RNA-seq datasets that there are
2844 sexually differentially methylated regions in the tilapia genome and that sex-
2845 biased gene expression is affected by CpG methylation level in the promoter
2846 regions, a knowledge gap still exists in what causes these sexually dimorphic
2847 CpG methylations. One possible mechanism may be through the dynamic
2848 regulation of CpG methylation landscape by androgen receptors. In human
2849 healthy prostate cells, it has been shown that hormone stimulation of androgen
2850 receptor can cause dynamic changes in the transcription rate and DNA
2851 methylation patterns at the AR target genes, TIPARP and SGK1. The target genes
2852 of androgen receptors in teleost are under-studied. Teleost specific whole genome
2853 duplication has led to two functional androgen receptor paralogs in teleost,
2854 namely AR α and AR β . It is shown that these two paralogs have different target
2855 genes in the genome through gain-of-function mutations. In future research, a
2856 CHIP-seq targeting the target gene of AR α , AR β , ESR1 and ESR2 in various
2857 tissues could be conducted in teleosts to understand the targets of various steroid
2858 hormone receptors in teleosts. However, prior to performing ChIP-seq,
2859 antibodies targeting these steroid hormone receptors should be raised and
2860 validated.

2861 By complementing WGBS with ChIP-seq, a more comprehensive cis-
2862 regulatory landscape can be inferred as it would include epigenetic regulatory
2863 regions that were inferred from regions with histone modifications. In *O.*
2864 niloticus, a ChIP-seq dataset targeting H3K4me3 derived from anal fin tissue is

2865 available. Our data will complement the existing available data for regulatory
2866 region analysis. In future works, to functionally analyse cis-regulatory regions,
2867 candidate regions mapped can be analysed with STARR-seq for quantitative
2868 analysis of enhancer regions. Due to higher interest in non-coding regions of
2869 DNA and their role in evolution, polymorphisms in CREs are thought to be
2870 responsible for causing adaptive phenotypic complexity through changes in gene
2871 expression and cell identity development. CREs are important genetic tools for
2872 in-depth analysis on the evolution of vertebrate transcriptional regulation.
2873 Although mutations in coding sequences that result in amino acid changes are
2874 often responsible for many phenotypic divergences, they are often expected to be
2875 pleiotropic, affecting more phenotypes and are more likely to be deleterious due
2876 to dis-functional amino acids. Phenotypic changes in CREs on the other hand,
2877 are thought to be more modular and tissue specific, as any changes in CRE
2878 regulation tend to produce tissue specific phenotypes. With improved knowledge
2879 on the coordinates of regulatory elements, more in-depth analysis can be done on
2880 genetic variants found on these non-coding regulatory elements, using data
2881 generated from QTL and GWAS experiments. Possible changes in expression
2882 due to polymorphisms in CREs may explain the phenotypic variation occurring
2883 due to differential expression.

2884

2885 **5.2 Muscle transcriptomes**

2886 **5.2.1 Sex-biased gene expression in tilapia skeletal muscle**

2887 In this thesis, I presented a transcriptomic analysis in the skeletal muscle of male
2888 and female in hybrid tilapia (*O. niloticus* X *O. mossambicus*) using the latest

reference genome, which allows us to understand the difference in sex-biased gene expressions between sexes. I obtained 19.1 million reads on average for each sample. This sequencing depth is similar to other transcriptomic analyses in teleosts^{230,231}. Here, I analyzed our RNA-seq data with the latest assembly, O_niloticus_UMD_NMBU²³². In this assembly, PacBio long reads were used to close the gaps as well as resolve complex repeats in previous assembly. This allowed us to map more than 92% of all reads onto the reference genome. Using the following statistical thresholds: false discovery rate corrected p-value < 0.05, and fold-change of at least two-fold, I detected 136 sex-biased genes in the skeletal muscle tissues with high confidence. This new reference genome provided better gene annotations and allowed for more accurate RNA-seq analysis. Based on our high R^2 value between RNA-seq expression level and qPCR expression value, I conclude that our RNA-seq DEG analysis is accurate.

2902

2903 **5.2.2 Pathways regulating cell growth and proliferations are sexually** 2904 **biased in expression**

From our results, I found that genes and pathways promoting growth are upregulated in males, while genes related to negative regulation of growth are upregulated in females. One of the pathways upregulated in males is the *MAPK* signaling pathway. *RASGRF1*, an upstream gene in the *MAPK* signaling pathway²³³, is highly expressed in male skeletal muscle tissue. In *RASGRF1* knock-out mice, somatic growth has been reported to be delayed. *RASGRF1*-deficient mice also showed hypoinsulinemia and low glucose tolerance due to a reduction of Beta-cells in the pancreatic islet, which produces insulin²³³. In

2913 another study of sexual size dimorphism in *C. semilaevis*, *MAPK* pathway was
2914 also found to be upregulated in females, which has faster growth rate compared
2915 to males⁶. One of the pathways upregulated in females, is the *FoxO* signaling
2916 pathway, which consists of a transcription factor family responsible for
2917 suppressing the expression of genes related to cell growth, proliferation and
2918 differentiation. One of its members, which I found to have a female-biased
2919 expression in tilapia skeletal muscle, is *FoxO4*. This gene has been shown to be
2920 an inhibitor of myogenic differentiation in muscle tissues, hence it is a negative
2921 regulator of somatic growth²³⁴. Another gene from the *FoxO* signaling pathway
2922 that is female-biased in expression is the insulin receptor substrate 2b (*IRS2b*).
2923 *IRS2b*^{-/-} mice have been shown to be 30% heavier in bodyweight at 32 weeks old,
2924 suggesting the role of *IRS2b* in regulating growth and energy homeostasis²³⁵.

2925 Our study provided novel insights into sex-biased genes in a somatic tissue,
2926 the skeletal muscle tissue. I showed that many genes (e.g. *RASGRF1*, *FoxO4* and
2927 *IRS2b*) are sex-biased in expression and may contribute to sexual size
2928 dimorphism, as these genes are key regulators of cell growth and proliferations.
2929 Given our data, I would like to suggest that positive regulators of growth (e.g.
2930 *RASGRF1*) are upregulated in male tilapia muscle tissue while negative regulator
2931 of cell growth and proliferations (e.g. *FoxO4* and *IRS2b*) are upregulated in
2932 females. However, I would like to highlight that these genes are not located on
2933 chromosomes known to be associated with the sex determining locus. Our results
2934 suggest that there may be other genetic factors in regulating sexual size
2935 dimorphism, besides sex determining genes. The mechanisms of achieving sex-
2936 biased gene expression in the presence of different sex chromosomes remain

2937 largely unknown in tilapia. In future works, it may be essential to design an
2938 experiment to investigate how sex hormones (e.g. Testosterone) or sex-
2939 determining genes (e.g. *AMH* and *DMRT1*) regulate somatic gene expression in
2940 fish. It could be also interesting to link sex determination, culture condition, DNA
2941 methylation and sexual dimorphism.

2942 **5.2.3 Sex-biased genes located in QTLs related to growth**

2943 I also identified three sex-biased candidate genes that are located within a QTL
2944 for higher growth rate in hybrid tilapia (*O. niloticus* x *O. mossambicus*)⁴⁴. One
2945 of the genes is *CADM3*, which is involved in cell to cell adhesion activity. The
2946 second gene is a novel gene, LOC100706127, with no known function. It has a
2947 highly conserved immunoglobulin domain, suggesting that it may be involved in
2948 immune function. The third sex-biased gene located in this QTL is a non-coding
2949 gene, termed LOC109195677. It is located within the intronic region of the gene
2950 LOC100706127. The function of this ncRNA is unknown. In future studies, these
2951 three genes may be candidate genes for identifying DNA markers for performing
2952 association studies related to growth in tilapia and for markers-assisted selection
2953 for growth.

2954 In addition, I also observed three sex-biased genes from chromosome LG1
2955 located within a QTL previously associated with sex phenotype in Mozambique
2956 tilapia. These genes are chitin synthase (*CHS-1*), phosphatidate phosphatase
2957 (*LPIN1*) and ras-specific guanine nucleotide-releasing factor 1 (*RASGRF1*).
2958 *CHS-1* are responsible for the synthesis of chitin molecules²³⁶. Morpholino
2959 knock-down of *CHS-1* in zebrafish resulted in unstable embryonic development
2960 in zebrafish larvae²³⁶. *LPIN1* is responsible for the final step in triglyceride

2961 synthesis and the expression of this gene is also required for adipocyte
2962 differentiation²¹⁵. *LPIN1* also plays a role in insulin sensitivity and regulation of
2963 metabolisms in cells²¹⁵. The functions of *RASGRF1* has been discussed in
2964 previous segments. Due to the close positions of these genes to the sex
2965 determining region in chromosome LG1, these genes may be linked to the sex
2966 determining region.

2967 Besides protein coding genes, I also identified five non-coding RNA genes in
2968 the list of sex-biased genes. These genes may play regulatory roles in fine tuning
2969 the expression levels of other genes via chromosomal inactivation, such as in the
2970 case of *Xist* which inactivates one of the X-chromosomes in female placental
2971 mammals²³⁷. However, the functions of these five non-coding RNAs are
2972 unknown. It is also interesting to note that although our libraries were constructed
2973 with oligo-dT enrichment methods, I was able to enrich some of the non-coding
2974 RNAs into our cDNA libraries. Studies have shown that a portion of known non-
2975 coding RNAs were known to have poly-A tails, hence it is possible for some non-
2976 coding RNAs to be enriched in oligo-dT enriched library^{238,239}. However, it
2977 should be pointed out that if one wishes to perform transcriptomic analysis
2978 targeted towards non-coding RNA, a ribosomal depletion method should be used
2979 instead²⁴⁰.

2980 **5.3 Brain transcriptome**

2981 **5.3.1 Use of microsatellite markers for molecular sexing**

2982 In this study, I presented the first transcriptomic analysis of sex-biased gene
2983 expression in tilapia brain tissue. I used microsatellite markers linked to sex

determining region to genotype male and female parents and juvenile fish. This allows us to distinguish the genetic sex of each sample even though the juvenile tilapia had not reached sexual maturation stage and visual sexing is not possible. Using this approach, I identified regions of no genetic recombination, also known as haplotype blocks. As the markers flanking the sex determining regions were co-segregated in all samples in a single haplotype block with no recombination detected in this region, I can confidently assign sex to our fish according to their microsatellite alleles.

Another crucial requirement for the success of this sequencing experiment is good dissecting methods. By decapitating the fish dorsal of the gill flaps and making careful incision around the orbital skull of the fish, I can keep the brain tissue intact, without contamination of tissues from other organs. As RNA-seq technology is highly sensitive, contamination of RNAs from other tissues may skew the detection power of sex-biased genes in our experiment.

5.3.2 Sex biased gene expression in the brain

I selected three crucial time points in early sex development in tilapia for my brain transcriptome project, which were 5 dph, 30 dph and 90 dph. At 5 dph, the sex determining mechanisms in tilapia are initiated²⁰⁹. At 30 dph, in female gonads, germ cell meiosis and oocyte formation are initiated^{37,209,210}. At 90 dph, in the male gonads, germ cell meiosis and spermatogenesis are initiated^{37,209,210}. Masculinization or feminization of the teleost brain may begin during the early

3007 stages of development, which may contribute towards sexual size dimorphism in
3008 tilapia. By sampling the brain tissue at these three crucial time points, I am able
3009 to detect the effect of sexual differentiation on somatic tissues such as the brain.

3010 My transcriptome data averages at 31.6 million reads per samples, with
3011 an average of 20.6 million reads mapped uniquely to the reference genome for
3012 each library. The high sequencing coverage and the use of up to four biological
3013 replicates per sex in our transcriptome allows us to increase the statistical power
3014 in detecting sex-biased genes. I defined sex-biased genes in our experiments as
3015 genes with sex-biased gene expression of 2-fold differences and above with false
3016 discovery rate corrected P-value of less than 0.05. In my results, I detected only
3017 moderate number of sex-biased genes at 5 dph and 30 dph in the brain
3018 transcriptome after taking into consideration false discovery rate. However, at 90
3019 dph, I observed high numbers of sex-biased genes in the brain. This may have
3020 been because at 90 dph, the gonads of tilapia are already mature and functioning.
3021 Masculinization or feminization of the tilapia brain may have already taken place
3022 as a result of exposure to sex steroid hormones. At 90 dph, sexually dimorphic
3023 behaviours such as courtship behaviour by male tilapia through the darkening of
3024 lateral skins and aggressive behaviour towards other males can already be
3025 observed in Mozambique tilapia. Also, the external genitalia of tilapia is already
3026 distinguishable at 90 dph, suggesting that sex maturation is already complete in
3027 both sexes. Hence, at 90 dph, sexual maturation may have altered the gene
3028 expression architecture in tilapia of both sexes, hence causing high number of
3029 sex-biased gene expression.

3030 I looked at the distribution of the sex-biased genes in our brain
3031 transcriptome datasets and found that the sex-biased genes are non-randomly
3032 distributed across the genome. This observation is similar to our muscle
3033 transcriptome datasets. A possible explanation to this could be trans-regulatory
3034 factors originating from sex chromosome region or originating from
3035 steroidogenic hormone receptors, which triggers the expression of sex-biased
3036 genes in different chromosomes. The teleost brain organ is a target of many
3037 steroidogenic hormones, such as testosterone and estrogen. In our dataset, I
3038 observed male-biased high expression of androgen receptor (*AR*) and estrogen
3039 receptor type 2 (*ESR2*) in 90 dph development stage, but not in earlier stages.

3040

3041 **5.3.3 Sex-biased genes in the brain are not randomly distributed in the** 3042 **genome**

3043 I looked at the distribution of the sex-biased genes in our brain transcriptome
3044 datasets and found that the sex-biased genes are non-randomly distributed across
3045 the genome. This observation is similar to our muscle transcriptome datasets. A
3046 possible explanation to this could be trans-regulatory factors originating from sex
3047 chromosome region or originating from steroidogenic hormone receptors, which
3048 triggers the expression of sex-biased genes in different chromosomes. The teleost
3049 brain organ is a target of many steroidogenic hormones, such as testosterone and
3050 estrogen. In teleosts, testosterone and estrogen exist in differing concentrations
3051 within the plasma serum. Testosterone and estrogen exert their functions in
3052 regulating target gene expression level via androgen receptor and estrogen
3053 receptor, respectively. In our dataset, I observed male-biased high expression of

3054 androgen receptor (*AR*) and estrogen receptor type 2 (*ESR2*) in 90 dph
3055 development stage, but not in earlier stages. These hormone receptors may bind
3056 to the cis-regulatory regions found in close proximity to target genes and trigger
3057 sex-biased expression. This may explain the reason why sex-biased genes in the
3058 tilapia brain are not randomly distributed in the genome.

3059 **5.3.4 Hypothalamic-Pituitary Axes are overrepresented with sex-biased**
3060 **genes.**

3061 Gene ontology results of my transcriptomic data suggests that the sex-biased
3062 genes in our transcriptome datasets are involved in biological process such as
3063 sensory system development, head development and nervous system
3064 development. These results are expected as I only dissected whole brain for RNA
3065 isolation, hence the majority of the genes expressed are in association with the
3066 central nervous system and the pituitary glands. In gene pathway analysis using
3067 PantherDB, I found that many pathways from the hypothalamic–pituitary–
3068 interrenal gland (HPI) axis, hypothalamic–pituitary–thyroid (HPT) axis and
3069 hypothalamic–pituitary–gonadal (HPG) axis are statistically overrepresented¹⁸⁴.
3070 In these axes, neuropeptide signals from the central nervous system are released
3071 into the pituitary lobes, which secrete neuroendocrine hormone signals that can
3072 activate processes in organs such as the adrenal glands, thyroid and gonadal
3073 tissues. All three axes have associated functions in regulating growth and
3074 reproduction. HPT exerts control over growth through the secretion of thyroid
3075 hormones in the thyroid glands. Thyroid hormone treatment using T2 and T3
3076 have been shown to stimulate growth in tilapia^{55,241}, while enhancing the
3077 expression of *IGF1* in the liver of tilapia⁵⁵.

3078 Besides the HPT axis, I also observed enrichment of sex-biased genes in
3079 the HPG axis. In the HPG axis, GnRH stimulates the secretion of Luteinizing
3080 Hormone (LH) and Follicle-stimulating Hormone (FSH) in the pituitary
3081 glands^{71,242-244}. LH and FSH can then stimulate the secretion of sex steroid
3082 hormones in the gonads which then provides positive feedback back to the
3083 hypothalamus²⁴²⁻²⁴⁶. Our observation of enrichment in sex-biased genes in the
3084 HPG axis is to be expected as male and female tilapia have a different rate of sex
3085 maturation as highlighted in gonadal dissection in other studies^{209,210}. One of the
3086 genes central to HPG axis is gonadotrophin releasing hormone 3 (GnRH3). In
3087 our sex-biased gene expression results, I found that GnRH3 has a male-biased
3088 expression. GnRH3 has been reported to be essential for nest building and
3089 aggressive behaviour between male tilapia, suggesting that GnRH3 is a potent
3090 neuromodulator of reproductive behaviours in male tilapia²⁴⁷. In the African
3091 cichlid *A. burtoni*, male to male visual threat exposure for three days significantly
3092 raised GnRH3 expression level, which further supports the role of GnRH3 in
3093 male reproductive behaviour²⁴⁸. GnRH3 ablation experiments in zebrafish
3094 resulted in sterile individuals and CRISPR/Cas9 GnRH3 knockout in zebrafish
3095 reveals no effect on reproductive performance in zebrafish¹⁶⁵. There have not
3096 been reports of GnRH3 regulating growth or sexual size dimorphism in other
3097 species, hence, the sex-biased expression of GnRH3 in tilapia brain may have
3098 been responsible for sexual dimorphism in behaviour instead of being related to
3099 sexual size dimorphism in tilapia.

3100 In the HPI axis, cortisol is induced by *ACTH* peptide in the adrenal cortex,
3101 which is a cleaved peptide of the protein precursor POMC¹⁶⁹. In teleost, the

3102 interrenal glands are the equivalent of adrenal cortex in mammals. The interrenal
3103 gland is a bilobe structure lateral to the notochord tissue and is composed of an
3104 epithelium made up of steroid producing cells and chromaffin cells. This gland
3105 is situated within the head kidney. In the interrenal gland, *ACTH* binds to
3106 melanocortin receptor 2 (*MC2R*) and promotes cortisol secretion. Cortisol is a
3107 well-studied hormone associated with stress and energy homeostasis^{57,169}. In
3108 Mozambique tilapia *in vivo* and *in vitro* experiments, cortisol has been shown to
3109 reduce the expression of *IGF-1* even in the presence of *GH* in hepatocyte cells
3110 found in the liver, where growth promoting *IGF-1* is expressed^{249,250}. Similar
3111 results were also reported in salmonids and rainbow trout⁵⁷. In our transcriptome
3112 data and qPCR data, I found consistent female-biased *POMC* expression level at
3113 30 dph, 90 dph and 120 dph in the brain. Besides inducing stress related hormones,
3114 *POMC* has also been reported to induce sexual dimorphism in energy
3115 homeostasis in mice¹⁶⁸ and diet-induced growth in zebrafish⁸³. These functions
3116 may be attributed to the anorexigenic function of *POMC* through its peptide
3117 subunits of α -*MSH* and its receptor MC4R, which is located in the paraventricular
3118 nucleus. As such, *POMC* is a suitable candidate gene in downstream functional
3119 characterization experiment. The ability of *POMC* to regulate somatic growth has
3120 also been reported and will be discussed in the following segment.

3121 **5.4 *POMC* as a candidate gene for sexual size dimorphism in teleost.**

3122 As tilapia showed female-biased *POMC* expression at 30 dph and 90 dph, I
3123 decided to generate a teleost loss-of-function model of *pomc* to study the function
3124 of *POMC* on somatic growth. As of the writing of this thesis, there has not been
3125 a *pomc* knockout mutant reported in a teleost model. On the ZFIN database, only

3126 morpholino knockdown of *POMC* in zebrafish larvae has been reported, which
3127 resulted in smaller melanophore size, similar to our F0 mutants²⁵¹. Loss-of-
3128 function in *POMC* has only been demonstrated in mice. Ideally, this loss-of-
3129 function analysis is to be performed in our species of interest, which is the tilapia.
3130 However, given the long reproductive cycle of tilapia, I decided to study the
3131 effect of loss-of-function in *POMC* in zebrafish for the following reasons. Firstly,
3132 zebrafish has a significantly shorter reproductive cycle and higher fecundity
3133 compared to tilapia. A pair of zebrafish at 120 dph can produce up to at least 100
3134 embryos per week. In contrast, a pair of Mozambique tilapia at 180 dph can only
3135 produce approximately 100 embryos every month. Secondly, zebrafish mating is
3136 more predictable and makes it a superior candidate in comparison with tilapia.
3137 Zebrafish can produce embryos after 1 hour of photoperiod under well-lit
3138 condition in the morning²⁵². Tilapia on the other hand, only reproduce once every
3139 month and the breeding timing of tilapia is not as consistent and predictable as
3140 zebrafish¹⁸. The only practical way to reliably obtain single cell stage tilapia
3141 embryo is via artificial insemination, by stripping male milt and female tilapia
3142 oocytes¹⁸. However, this method is not feasible for my project, as females can
3143 only be stripped once a month and it causes significant amount of physiological
3144 stress to the brooders. A large stock of matured and well-developed brooders,
3145 estimated to be around 20 pairs, will be needed to allow weekly artificial
3146 insemination of tilapia oocytes for CRISPR/Cas9 experiments, which is a
3147 logistical challenge for our marine facility. Thirdly, zebrafish embryos are
3148 transparent and are easier to be penetrated by glass capillary micropipettes in
3149 comparison to tilapia embryos. Tilapia embryos are opaque yellow in colour and
3150 are harder to penetrate with glass microcapillary pipettes. Also, the single cell

3151 animal pole is clearly visible in zebrafish embryo, but not in the case of tilapia
3152 embryo due to the yellow opaque embryo. Finally, the availability of genetic
3153 toolkits and other transgenic reporter lines in zebrafish makes zebrafish a better
3154 candidate in comparison with tilapia. Various mutant and morpholino
3155 knockdown lines related to melanocortin system such as *mc4r* and *agrp* have
3156 been established in zebrafish, allowing detailed downstream mechanism studies
3157 in future works.

3158 **5.4.1 *pomc* mutants have smaller melanophore distribution compared to** 3159 **wildtype**

3160 In our mutants, I observed that the melanophore distribution is smaller than in
3161 wildtype fish. This phenotype may be related to the function of *POMC* in
3162 regulating camouflage behavior²⁵¹. The corticotrophin-releasing factor (*CRF*)
3163 and *POMC* pathway are critical regulators in camouflage behavior²⁵¹. One of the
3164 peptides produced by the polypeptide cleavage of *POMC* is the hormone α -MSH,
3165 which can promote melanosome dispersal in cultured melanocytes and in larval
3166 zebrafish. Morpholino knock down of *POMCa* in zebrafish larvae have been
3167 shown to suppress the ability of melanosome to disperse when morpholino
3168 injected larva is placed in dark background location²⁵¹. The effect of *pomc*
3169 knockout mutation on the coat color of mouse has also been demonstrated.
3170 *POMC* null mice have a lighter ventral coat color in comparison with age-
3171 matched wildtype mice. Although in my thesis, the phenotype of reduced
3172 melanophore dispersal is only observed in mosaic F0 fishes, I hypothesize that in
3173 F2 homozygous *pomc* mutants, the inability of melanosome to disperse will be
3174 more obvious and ubiquitous.

3175 **5.4.2 *pomc* mutants exhibit higher growth rate**

3176 In my mutant crosses, I observed that *pomc* mutants have a higher growth rate
3177 than wildtype fish. *POMC* has been reported to be central to the regulation of
3178 energy homeostasis in vertebrates. In rodents²¹⁸, dogs²¹⁷ and humans²¹⁸, *POMC*
3179 null mutants have been reported to result in moderately enhanced linear somatic
3180 growth and/or obesity. In dogs, an 18-bp deletion in *POMC* has led to increased
3181 susceptibility to obesity, even in heterozygote dogs²¹⁷. In fish, somatic growth is
3182 regulated by *GH* released from the pituitary glands and gonads^{52,74,76,77,79}.
3183 Positive stimulants of the *GH* include the growth hormone releasing hormone
3184 (*GHRH*) while somatostatin (*SST*) inhibits the release of GH from the pituitary
3185 glands⁷³. Somatostatin producing cells express melanocortin 4 receptors (*MC4R*),
3186 which bind to alpha melanocyte stimulating hormone (α *MSH*), which is a product
3187 of *POMC*¹⁶⁸. Activation of *MC4R* by α *MSH* leads to the secretion of *SST* which
3188 inhibits the release of *GH* by the pituitary glands and leads to slow linear growth
3189 rate¹⁶⁸.

3190 In my thesis, I presented phenotypic data based on heterozygous F1
3191 mutants, which may not truly reflect the effect of homozygous *POMC* mutation
3192 on growth as compared to a stable homozygous *pomc*^{-/-} line. In future works,
3193 when homozygous lines have been established from our *pomc* mutants, I intend
3194 to repeat this experiment using larger biological replicates to confirm the results
3195 presented in this thesis.

3196 **5.4.3 *pomc* mutants exhibit short feeding latency behaviour in**
3197 **comparison with wildtype fishes**

3198 In my analyses of *pomc* fish feeding behaviour, I found that *pomc* zebrafish have
3199 a very short feeding latency time. This phenotype may be attributed to the
3200 function of *POMC* in controlling feeding behaviour¹⁶⁸. The melanocortin system
3201 has been shown to regulate feeding in mice²¹⁸. *MC4R*, one of a family of G-
3202 protein-coupled, 7-transmembrane domain receptors, was detected in
3203 hypothalamic regions which are involved in regulating feeding behaviour²⁵³.
3204 These regions in mice includes the arcuate (ARC), paraventricular (PVN), and
3205 hypothalamic dorsomedial (DMH) and ventromedial (VMH) nuclei^{172,218}.
3206 Agonists that showed binding and activation of *MC4R* include α -MSH and β -
3207 MSH, which are both derived from the cleavage of POMC polypeptide¹⁶⁸. Out of
3208 these two ligands, β -MSH showed stronger and robust inhibition of feeding
3209 through binding to *MC4R*¹⁶⁶. In mice, under food restriction control, β -MSH has
3210 been found to accumulate inside the VMH, DMH and ARC while under free
3211 feeding condition, VMH, DMH and ARC in mouse show low accumulation of β -
3212 MSH¹⁶⁸. It was proposed in mice that the accumulation of β -MSH reflects a lack
3213 of peptide release to the surrounding nerve cells containing *MC4R*, which leads
3214 to increased hunger in mouse and also increased expression of *MC4R* to increase
3215 the sensitivity of *MC4R* towards β -MSH¹⁷⁰. In dogs, an 18-bp deletion in POMC
3216 also leads to behaviours of increased food motivations²¹⁷. In *pomc* zebrafish,
3217 target genome editing sites are located on the genomic region encoding β -MSH
3218 in *POMC* gene. The robust feeding behaviour observed in *pomc* fish may be
3219 attributed to the lack of β -MSH in *pomc* fish, hence altering the feeding

3220 regulation mechanism in *pomc* fish. Thus, it is possible that the *pomc* fish are
3221 constantly in a state of hunger in comparison with wildtype fish and showed the
3222 short latency to feeding behaviour as reported in my thesis.

3223

3224 **5.4.4 Transcription of *POMC* in tilapia may be induced by estrogen**

3225 In my thesis, I showed that two predicted estrogen response elements located at
3226 1379 bp and 2388 bp upstream of *POMC* transcriptional start site can be induced
3227 by estrogen treatment through a luciferase reporter assay. The two predicted
3228 EREs have more than 85% similarity to DNA binding site motifs for estrogen
3229 receptor beta in mice. Based on our results, estrogen may regulate the expression
3230 of *POMC* in tilapia. In tilapia, estrogen levels are higher in female tilapia in
3231 comparison with male tilapia²⁵⁴. The main active form of estrogen in tilapia is
3232 the 17 β -estradiol, also known as β -estradiol. During breeding cycles, 17 β -
3233 estradiol level in female tilapia serum fluctuates at different stages of an ovarian
3234 cycle²⁵⁴. Being a mouth brooding species, the serum level of 17 β -estradiol in
3235 tilapia was shown to be different in mouth brooding females in comparison to
3236 non-mouth brooding females²⁵⁴. The concentration of 17 β -estradiol has been
3237 shown to peak at 15 days after spawning in mouth brooding female tilapias and
3238 remains at high level until mouth brooding ceases²⁵⁴. Similar observations have
3239 also been reported in other closely related mouth brooding cichlid fish species
3240 such as *A. burtoni*^{164,255,256}. Hence, it may be possible that the high estrogen levels
3241 in mouth brooding females promote the expression of *POMC* and suppresses
3242 feeding behaviour in sexually matured female tilapia until mouth brooding ceases.

3243 In teleosts, the two main sources of estrogen are the ovary and the brain¹⁶².
3244 Upon sexual maturation, GnRH1 will stimulate the secretion of FSH in the
3245 female pituitary glands¹⁶³. Secretion of FSH will stimulate the synthesis of
3246 estrogen by granulosa cells in the ovary²⁵⁷. The estrogen is then cycled to the
3247 whole body through the blood circulation system. This mode of estrogen
3248 production is dependent on neuroendocrine signals from the brain and pituitary
3249 glands, namely GnRH1 from the hypothalamus and FSH from the anterior
3250 pituitary lobe. The temporal control of this hormone synthesis is also cyclical and
3251 follows a monthly ovulation cycle. The other source of estrogen is from the
3252 brain¹⁶². In teleost, the aromatase gene (*CYP19A1*) is highly active in the brain¹⁶².
3253 Aromatase converts testosterone into estrogen in the brain¹⁶². In tilapia, female
3254 brain has a higher aromatase activity in comparison to male²⁵⁸. It may be possible
3255 that aromatase triggers the increased concentration of estrogen in the female brain
3256 tissue, leading to higher *POMC* expression in the female brain. From my RNA-
3257 seq data, aromatase has a higher expression in female at 30 dph and 90 dph but
3258 the differences from male are not statistically significant ($P = 0.33$ and $P = 0.61$,
3259 respectively). The differences are also less than 2-fold, hence the statistical
3260 significance may have been penalised due to that. In another study, XX female
3261 tilapia has been shown to have higher aromatase activity in comparison with XY
3262 male tilapia²⁵⁸. The difference in aromatase activity was also less than 2-fold in
3263 that study. In future works, more detailed studies should be conducted on the
3264 aromatase activity difference between male and female tilapia and its role in
3265 feminization of the tilapia brain.

3266 It is interesting to note that in mice, female-biased expression of *POMC*
3267 is also prevalent¹⁷². In mice, *TAp63* protein has been shown to cause sexually
3268 dimorphic expression of *POMC*, hence leading to differential regulation of
3269 energy homeostasis in male and female mice. *TAp63* is a transcriptionally active
3270 variant of *p63* found in mice and has been found to be a master regulator in lipid
3271 and glucose metabolism. In our RNA-seq datasets, the *TAp63* orthologue in
3272 tilapia did not show sexually dimorphic expression. This suggests that some other
3273 transcription factors may be involved in regulating the sex-biased expression of
3274 *POMC* in tilapia.

3275 **5.4.5 Proposed model of *POMC* mechanism in tilapia sexual size**
3276 **dimorphism**

3277 In my thesis, through RNA-seq data and qPCR verifications, I have shown that
3278 *POMC* expression is female biased from 30 dph up to 120 dph in tilapia. Using
3279 luciferase assay, I showed that two EREs located upstream of *POMC*
3280 transcriptional start sites are sensitive to induction by estrogen treatment, hence
3281 explaining the possible mechanism through which female-biased *POMC*
3282 expression is achieved. Using loss-of-function functional analyses of *POMC* in
3283 zebrafish, I found that lack of *POMC* expression led to increase in somatic growth
3284 and increased sensitivity towards feeding in zebrafish.

Figure 41. Proposed mechanism of *POMC* in regulating sexual size dimorphism in tilapia. Pink arrows denote *POMC* model in female and blue arrows denote *POMC* model in male. Higher level of estrogen in female tilapia led to upregulation of *POMC*, resulting in slow growth rate in females. Low level of estrogen in male tilapia led to a basal level of *POMC* transcription in male tilapia, resulting in faster growth rate in male tilapia. The mechanisms through which *POMC* influences feeding behaviour in tilapia is not studied in this thesis, although *pomc* mutants in zebrafish showed increased feeding behaviour and increased somatic growth. *POMC* may also extend its mechanisms to allow for mouth brooding behaviours in tilapia.

Hence, it may be inferred that low *POMC* expression in male tilapia may have increased growth rate in male tilapia, leading to male-biased sexual size dimorphism in tilapia. On the other hand, consistently high expression of *POMC* in female tilapia due to estrogen influence may have inhibited the release of growth hormones in the somatotropes of female tilapia, leading to slower growth rate in female tilapia. High level of *POMC* through induction by estrogen hormone in females may also be an evolutionary feature of mouth brooding behaviour to suppress feeding behaviour during mouth brooding seasons.

3305 However, further works are required to test the relationship between mouth
3306 brooding behaviours and *POMC* expression. Taken together, *POMC* is likely one
3307 of the causative genes for sexual size dimorphism in tilapia. A summary cartoon
3308 model is shown in Figure 41, summarizing the proposed mechanisms of *POMC*
3309 on sexual size dimorphism in tilapia.

3310

3311 **5.4.6 Future works on *POMC* and its possible applications in aquaculture**

3312 In future works, CRISPR/Cas9 genome editing of *POMC* in tilapia can be
3313 explored. In this thesis, I showed that *POMC* knockout in zebrafish can lead to
3314 increased somatic growth and higher sensitivity to feeding. However, the effect
3315 of *POMC* knockout in tilapia is unknown. There are two knockout strategies that
3316 need to be explored. Firstly, by removing the two EREs located upstream of
3317 tilapia *POMC*, I may investigate if removal of EREs can bring the expression
3318 level of *POMC* back to being equal in both sexes and also modify its effect on
3319 the mouth brooding behaviour in female tilapia. Secondly, I may target the
3320 coding sequences of *POMC* in tilapia to induce a frame-shift or deletion in amino
3321 acid sequences and study the effect of loss-of-function in *POMC* in tilapia. Based
3322 on our data from zebrafish *POMC* loss-of-functions, it is hypothesized that
3323 *POMC* loss-of-function in tilapia will lead to increased somatic growth. It is also
3324 crucial to investigate if *POMC* loss-of-function may reduce or remove sexual size
3325 dimorphism in tilapia.

3326 Another possible experiment would be to orally feed ESR inhibitors to
3327 female tilapia and check for the changes in *POMC* expression levels in female

3328 tilapia. Examples of ESR inhibitors include Tamoxifen²⁵⁹ and ICI 182,780²⁶⁰.
3329 While I showed that *POMC* may have been induced by estrogen treatment in
3330 luciferase assays, the *in vivo* effect of *POMC* induction by estrogen has not yet
3331 been investigated. There are reports of using Tamoxifen to induce sex reversal in
3332 tilapia, where oral administration of Tamoxifen from hatchling to 60 dph resulted
3333 in sex reversal to an 80% male and 1% intersex population²⁶¹. This suggests that
3334 Tamoxifen is effective in inhibiting teleost ESR. Other potential side effects such
3335 as mouth brooding behavioural changes may also be investigated as well.
3336 However, this experiment may lead to off-target effects as the genome-wide
3337 targets of ESRs are affected by ESR inhibitor treatment, for example, sex-
3338 reversal as reported by other groups²⁶¹.

3339 Although in this thesis, I showed that *POMC* knockout has a positive
3340 effect on the growth rate of zebrafish as well as increased sensitivity to feeding,
3341 the effect of *POMC* knockout on feed conversion ratio (FCR) remains unknown.
3342 FCR refers to the efficiency of how an aquaculture species converts feeds into
3343 desired economic traits (e.g. bodyweight, size or length). A low FCR is highly
3344 desirable in aquaculture, as it implies that less fish feed is required for aquaculture
3345 production. In *POMC* knockout zebrafish, the fish were shown to be more
3346 sensitive to feeding by showing a significantly shorter latency to feeding.
3347 However, our experiment set up does not allow a detailed investigation into the
3348 feed conversion ratio in *POMC* knockout zebrafish. In future research, the effect
3349 of *POMC* deletion and feed conversion ratio in *pomc*^{-/-} zebrafish should be
3350 investigated to explore the promise of using genome editing in aquaculture fish
3351 to improve feed conversion ratio in the aquaculture industry.

3352 In the future, the *pomc* zebrafish line may be used as a model to study the
3353 effects of *POMC* loss-of-function in teleost. It will be interesting to study the
3354 effect of *pomc*^{-/-} on the development of young embryos, as our analysis and
3355 reports by others showed that POMC has high expression level during early
3356 development period. It is also important to look into the function of *pomcb* in the
3357 event of *pomca* deletion. Recent research has suggested that a loss-of-function
3358 mutation in a gene will lead to activation of its paralog within the same genome
3359 through a novel mRNA degradation pathway²⁶². By knocking out *POMCa* in a
3360 *pomc* mutant, increased transcription of *POMCb* may occur to compensate for
3361 the loss of *POMCa*.

3362 **5.5 Single nucleotide polymorphism markers for marker-assisted** 3363 **selection.**

3364 In this report, I identified three SNPs located on the promoter region of
3365 RASGRF1 in chromosome LG1 which are associated with bodyweight in tilapia.
3366 I also used brain transcriptome datasets to call SNPs and indels located on the
3367 transcribed regions of the tilapia genome. Some of the SNPs located at
3368 chromosome LG1 showed association with sex phenotype, which is consistent
3369 with previous reports regarding the sex determining region in Mozambique
3370 tilapia^{39,40}. These markers may be used as a resource for future market-assisted
3371 selection program in tilapia, especially when selecting traits related to growth
3372 and/or sex. However, it should be noted that some of these SNPs may have been
3373 fixed or show no polymorphisms in other populations.

3374 In this thesis, I also used transcriptome datasets to call for SNPs located in
3375 transcribed regions. This approach has different challenges compared to calling

3376 of SNPs from genome sequencing data, especially in regard to splicing events.
3377 Splicing events on RNA-seq datasets may lead to calling of false positive SNPs
3378 or indels, due to the many types of isoforms that may be formed from each gene.
3379 To overcome this problem, I employed STAR-aligner's two-pass alignment
3380 protocol, which is specifically designed to more accurately assign RNA-seq reads
3381 to their original chromosomal position by flagging splice junctions during
3382 mapping and re-mapping process of the two-pass alignment process. Overhangs
3383 located at exon-exon junctions may also cause false positive in calling SNPs from
3384 RNA-seq datasets, hence, GATK's SplitNCigar protocol was employed to clip
3385 away these overhangs to prevent false positive SNPs from being called. Finally,
3386 I also employed a high mapping quality (Q30 and above) and sequencing depth
3387 (20X) requirement to compensate for the uneven distribution of RNA-seq reads
3388 on a genome as compared to whole genome resequencing datasets.

3389 In this thesis, I also reported association of sex phenotype with a cluster or
3390 SNPs from chromosome LG1 in Mozambique tilapia. This result is consistent
3391 with earlier investigation into the sex-determining locus in Mozambique tilapia,
3392 which showed that the sex determining locus is located in LG1^{39,44}. However, it
3393 should be noted that in my association study, only 16 individuals were used as
3394 some samples did not meet the coverage requirement for SNPs calling. Due to
3395 the small sample size in this association study, more samples from different
3396 populations should be used to verify the effectiveness of these SNP markers.
3397 Subsequently, these markers can be used for identification of sex identity in
3398 Mozambique tilapia. One key challenge in Mozambique tilapia aquaculture is the

3399 lack of genetic markers for identification of sex, which impedes the development
3400 of monosex Mozambique tilapia. Using the SNPs markers identified in this

3401

3402 **5.6 Sex-biased genes in tilapia are under relaxed purifying selection.**

3403 Sexual size dimorphism is a form of phenotypic plasticity. At the molecular level,
3404 phenotypic plasticity gives rise to sexually dimorphic traits through the
3405 differential expression of genes among sexes. Sex-biased gene expression may
3406 have been an indirect consequence of sexual differentiation. Many sex-biased
3407 genes contribute towards fitness in sexual selection and signs of adaptive changes
3408 can be found in sex-biased genes. One such sign is the non-synonymous to
3409 synonymous mutation rates of protein coding sequences, also known as dN/dS .
3410 In this thesis, I calculated the non-synonymous to synonymous mutation rate in
3411 tilapia using coding sequences of tilapia and other closely related cichlid species.
3412 I found that in both skeletal muscle and brain tissue, sex-biased genes showed
3413 higher dN/dS in comparison with unbiased genes.

3414 Reports investigating sex-biased gene expression in other species such as
3415 *Drosophila*^{111,113,114,116}, zebrafish, guppy⁵⁸, chicken⁹³ and brown algae have
3416 shown that sex-biased genes exhibit higher dN/dS levels in comparison with
3417 unbiased genes. Hence, our results are consistent with reports from other species.
3418 However, it is important to note that although several of the sex-biased genes
3419 showed possible signs of positive selection ($dN/dS \geq 1$), the majority of sex-biased
3420 genes have a dN/dS level of less than 1. This suggests that the majority of sex-
3421 biased genes are under relaxed purifying selection while non-biased genes are
3422 under stronger purifying selection, as shown by the higher dN/dS values of sex-

3423 biased genes compared to the lower dN/dS values of non-biased genes. Although
3424 I showed that sex-biased genes in tilapia are under relaxed purifying selection,
3425 based on our data, it is still not clear if the elevated dN/dS is due to natural
3426 selection, sexual selection, relaxed constraint or genetic drift. One possible future
3427 direction is to test whether these mutations are caused by random mutations or
3428 non-random selection force such as natural selection or sexual selection. To do
3429 so, we first need a data set with high SNP density. Then, Tajima's D can be
3430 applied to test if the mutations are random (genetic drift) or under a direction-
3431 specific selection force (natural or sexual selection).

3432 **6 Concluding remarks.**

3433 Sexual size dimorphism is a complex quantitative trait controlled by multiple
3434 genes located on different chromosomal positions. In this thesis, I reported the
3435 transcriptional and epigenetic architecture of sexual dimorphism in tilapia. The
3436 first single base resolution methylome from tilapia is reported in this thesis. I
3437 found extensive differences between male and female tilapia methylomes,
3438 revealing the dynamic nature of CpG methylation in sexually dimorphic
3439 manners. I showed that the majority of cytosines in CpG context were largely
3440 methylated and CpG methylation in the promoter regions of tilapia were
3441 associated with suppressed gene expression. I also quantified the CpG
3442 methylation level in repeat elements in the tilapia genome, showing that most
3443 of the repeat elements, such as DNA transposons, LINEs, SINEs and LTRs,
3444 were highly methylated. I mapped UMRs and LMRs on the Nile tilapia
3445 reference genome (oreNil2), revealing a sexually dimorphic cis-regulatory
3446 landscape in the tilapia genome. I highlighted a sexually dimorphic regulatory

3447 landscape in tilapia, which may be involved in sexual dimorphism in tilapia. I
3448 also inferred different active transcriptional factors on CREs in male and
3449 female tissues, which may be used for future candidate CREs functional
3450 characterization.

3451 On the sexually dimorphic transcriptional architecture in tilapia somatic
3452 tissues, I reported sex-biased genes in skeletal and brain tissue of tilapia at
3453 different developmental stages. The gene ontologies and pathways related to
3454 these sex-biased genes have been described in detail. Many of these sex-biased
3455 genes are evolving at a faster pace compared to unbiased genes, as revealed by
3456 non-synonymous to synonymous substitution analysis. I also showed that
3457 neuroendocrine genes such as *POMC* and *GnRH3* are sex-biased in expression
3458 in the tilapia brain. I selected a candidate gene, *POMC*, and generated a loss-
3459 of-function knockout line in zebrafish as a model to study the role of *POMC*
3460 in regulating somatic growth in teleost. Here, I showed that in comparison with
3461 wildtype zebrafish, *pomc* mutants achieves a faster growth rate and a low
3462 feeding latency. I also tested a putative enhancer containing two EREs at 1.5
3463 kb upstream of tilapia *POMC* gene using luciferase assay and estrogen
3464 treatment and showed that the two EREs were sensitive to estrogen treatment
3465 and can induce a higher level of luciferase activity. These results suggest that
3466 the higher *POMC* expression in female tilapia may have been triggered by
3467 higher concentration of estrogen hormone in female tilapia brain through the
3468 two EREs tested, leading to slower growth rate in female tilapia.

3469 Finally, I showed that three SNPs located on the promoter regions of
3470 *RASGRF1* are associated with bodyweight. This set of markers may be used

3471 for marker-assisted selections for traits relating to growth. I also called a set of
3472 high-quality SNPs from the brain transcriptome and identified a cluster of
3473 SNPs located on LG1 which may be associated with sex phenotype in
3474 Mozambique tilapia. These SNPs may be used for molecular sexing in future
3475 works.

3476 Although in this thesis, I have showed that *pomc* may be functional in
3477 regulating growth rate differences between male and female tilapia, the effect
3478 of *pomc* on female brooding behaviours have not been studied. This may be
3479 the focus of future research endeavour on sexual dimorphism in tilapia.
3480 Besides *pomc*, other neuroendocrine peptides such as neuropeptide Y and
3481 GnRH should be examined as well as neuronal cells expressing these peptides
3482 are tightly linked in a clusters within the hypothalamus and the adjacent
3483 pituitary. Besides that, the targets of various steroid receptors may be
3484 examined in the future through the use of ChIP-seq techniques. In vertebrates,
3485 sex-biased genes may be driven by steroid receptors. Hence, it would be
3486 interesting to map out possible functional binding sites for steroid receptors in
3487 tilapia for the study of sexual dimorphisms.

3488

3489

3490 7 References

- 3491 1 Fairbairn, D. J., Blanckenhorn, W. U. & Székely, T. *Sex, size and gender roles: evolutionary studies of sexual size dimorphism*. (Oxford University Press, 2007).
- 3492 2 Parker, G. A. The evolution of sexual size dimorphism in fish. *Journal of Fish Biology* **41**,
- 3493 1-20 (1992).
- 3494 3 Isles, A. Sexual dimorphism the genetic way. *Trends in Genetics* **18**, 444 (2002).
- 3495 4 Mank, J. E. Sex Chromosomes and the Evolution of Sexual Dimorphism: Lessons from
- 3496 the Genome. *The American Naturalist* **173**, 141-150 (2009).
- 3497 5 Morita, K. & Tsuboi, J. Sexual size dimorphism in a landlocked Pacific salmon in relation
- 3498 to breeding habitat features. *Evolutionary Ecology* **31**, 653-661 (2017).
- 3499 6 Wang, N., Wang, R. K., Wang, R. Q. & Chen, S. L. Transcriptomics analysis revealing
- 3500 candidate networks and genes for the body size sexual dimorphism of Chinese tongue
- 3501 sole (*Cynoglossus semilaevis*). *Functional & Integrative Genomics* **18**, 327-339 (2018).
- 3502 7 Hedrick, A. V. & Temeles, E. J. The evolution of sexual dimorphism in animals:
- 3503 Hypotheses and tests. *Trends in Ecology & Evolution* **4**, 136-138 (1989).
- 3504 8 Ralls, K. & Mesnick, S. in *Encyclopedia of Marine Mammals* (eds William F. Perrin,
- 3505 Bernd Würsig, & J. G. M. Thewissen) Ch. Sexual dimorphism, 1005-1011 (Academic
- 3506 Press, 2009).
- 3507 9 Williams, T. M. & Carroll, S. B. Genetic and molecular insights into the development
- 3508 and evolution of sexual dimorphism. *Nature Reviews Genetics* **10**, 797 (2009).
- 3509 10 Fairbairn, D. J., Blanckenhorn, W. U. & Székely, T. *Sex, Size, and Gender Roles: Evolutionary Studies of Sexual Size Dimorphism*. (Oxford University Press, 2007).
- 3510 11 Chervinski, J. Sexual dimorphism in tilapia. *Nature* **208**, 703 (1965).
- 3511 12 Lind, C. E. *et al.* Differences in sexual size dimorphism among farmed tilapia species and
- 3512 strains undergoing genetic improvement for body weight. *Aquaculture Reports* **1**, 20-
- 3513 27 (2015).
- 3514 13 Schütz, D. & Taborsky, M. Giant males or dwarf females: what determines the extreme
- 3515 sexual size dimorphism in *Lamprologus callipterus*? *Journal of Fish Biology* **57**, 1254-
- 3516 1265 (2000).
- 3517 14 Mizzau, T. W., Garner, S. R., Marklevitz, S. A. C., Thompson, G. J. & Morbey, Y. E. A
- 3518 genetic test of sexual size dimorphism in pre-emergent chinook salmon. *PLOS One* **8**,
- 3519 e78421-e78421 (2013).
- 3520 15 Weir, L. K., Kindsvater, H. K., Young, K. A. & Reynolds, J. D. Sneaker Males Affect Fighter
- 3521 Male Body Size and Sexual Size Dimorphism in Salmon. *The American Naturalist* **188**,
- 3522 264-271 (2016).
- 3523 16 Wang, P. *et al.* Sexually dimorphic gene expression associated with growth and
- 3524 reproduction of Tongue Sole (*Cynoglossus semilaevis*) revealed by brain transcriptome
- 3525 analysis. *International Journal of Molecular Sciences* **17** (2016).
- 3526 17 Barneche, D. R., Robertson, D. R., White, C. R. & Marshall, D. J. Fish reproductive-energy
- 3527 output increases disproportionately with body size. *Science* **360**, 642 (2018).
- 3528 18 Webster, C. D. & Lim, C. *Tilapia: biology, culture, and nutrition*. (CRC Press, 2006).
- 3529 19 Klett, V. & Meyer, A. What, if Anything, is a Tilapia?—Mitochondrial ND2 Phylogeny of
- 3530 *Tilapiines* and the Evolution of Parental Care Systems in the African Cichlid Fishes.
- 3531 *Molecular Biology and Evolution* **19**, 865-883 (2002).
- 3532 20 FAO. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable
- 3533 development goals. (Food and Agriculture Organization of the United Nations, 2018).
- 3534 21 FAO. *Fisheries and aquaculture software. FishStat Plus - Universal software for fishery*
- 3535 *statistical time series.*, 2019).
- 3536 22 D'Amato, M. E. *et al.* Hybridization and phylogeography of the Mozambique tilapia
- 3537 *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and
- 3538 microsatellite DNA genotyping. *Conservation Genetics* **8**, 475-488 (2007).
- 3539
- 3540

- 3541 23 Ponzoni, R. W. *et al.* Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with
3542 special reference to the work conducted by the WorldFish Center with the GIFT strain.
3543 *Reviews in Aquaculture* **3**, 27-41 (2011).
- 3544 24 Oliveira, R. F. & Almada, V. C. Sexual dimorphism and allometry of external morphology
3545 in *Oreochromis mossambicus*. *Journal of Fish Biology* **46**, 1055-1064 (1995).
- 3546 25 Scott, A. G., Penman, D. J., Beardmore, J. A. & Skibinski, D. O. F. The 'YY' supermale in
3547 *Oreochromis niloticus* (L.) and its potential in aquaculture. *Aquaculture* **78**, 237-251
3548 (1989).
- 3549 26 El-Greisy, Z. A. & El-Gamal, A. E. Monosex production of tilapia, *Oreochromis niloticus*
3550 using different doses of 17 α -methyltestosterone with respect to the degree of sex
3551 stability after one year of treatment. *The Egyptian Journal of Aquatic Research* **38**, 59-
3552 66 (2012).
- 3553 27 Jensen, G. L. & Shelton, W. L. Effects of estrogens on *Tilapia aurea*: implications for
3554 production of monosex genetic male tilapia. *Aquaculture* **16**, 233-242 (1979).
- 3555 28 Dunham, R. A. Production and use of monosex or sterile fishes in aquaculture. *Reviews*
3556 *in Aquatic Sciences* **2**, 1-17 (1990).
- 3557 29 Dan, N. C. & Little, D. C. The culture performance of monosex and mixed-sex new-
3558 season and overwintered fry in three strains of Nile tilapia (*Oreochromis niloticus*) in
3559 northern Vietnam. *Aquaculture* **184**, 221-231 (2000).
- 3560 30 Beardmore, J. A., Mair, G. C. & Lewis, R. I. in *Reproductive Biotechnology in Finfish*
3561 *Aquaculture* 283-301 (2001).
- 3562 31 Directive, A. C. 96/22/EC concerning the prohibition on the use in stockfarming of
3563 certain substances having a hormonal or thyrostatic action and of β -agonist. (Directive
3564 2003/74/EC, Brussels, 2003).
- 3565 32 Hopkins, K. D., Shelton, W. L. & Engle, C. R. Estrogen sex-reversal of *Tilapia aurea*.
3566 *Aquaculture* **18**, 263-268 (1979).
- 3567 33 Lee, B. Y., Penman, D. J. & Kocher, T. D. Identification of a sex-determining region in
3568 Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* **34**,
3569 379-383 (2003).
- 3570 34 Shirak, A. *et al.* Amh and Dmrta2 genes map to tilapia (*Oreochromis spp.*) linkage group
3571 23 within quantitative trait locus regions for sex determination. *Genetics* **174**, 1573-
3572 1581 (2006).
- 3573 35 Eshel, O. *et al.* Fine-mapping of a locus on linkage group 23 for sex determination in
3574 Nile tilapia (*Oreochromis niloticus*). *Animal Genetics* **42**, 222-224 (2011).
- 3575 36 Palaikostas, C. *et al.* A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*).
3576 *BMC Genomics* **16**, 171 (2015).
- 3577 37 Li, M. *et al.* A tandem duplicate of anti-Müllerian hormone with a missense SNP on the
3578 Y Chromosome Is Essential for Male Sex Determination in Nile Tilapia, *Oreochromis*
3579 *niloticus*. *PLOS Genetics* **11**, e1005678 (2015).
- 3580 38 Campos-Ramos, R., Harvey, S. C., McAndrew, B. J. & Penman, D. J. An investigation of
3581 sex determination in the Mozambique tilapia, *Oreochromis mossambicus*, using
3582 synaptonemal complex analysis, FISH, sex reversal and gynogenesis. *Aquaculture* **221**,
3583 125-140 (2003).
- 3584 39 Liu, F. *et al.* A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis*
3585 *mossambicus* x *Oreochromis spp.*) and mapping of sex-determining loci. *BMC Genomics*
3586 **14**, 58 (2013).
- 3587 40 Liu, F. *et al.* A genome scan revealed significant associations of growth traits with a
3588 major QTL and GHR2 in tilapia. *Scientific Reports* **4**, 7256 (2014).
- 3589 41 Rutten, M. J. M., Komen, H., Deerenberg, R. M., Siwek, M. & Bovenhuis, H. Genetic
3590 characterization of four strains of Nile tilapia (*Oreochromis niloticus* L.) using
3591 microsatellite markers. *Animal Genetics* **35**, 93-97 (2004).
- 3592 42 Fessehay, Y. *et al.* Mating systems and male reproductive success in Nile tilapia
3593 (*Oreochromis niloticus*) in breeding hapas: a microsatellite analysis. *Aquaculture* **256**,
3594 148-158 (2006).

3595 43 Hassanien, H. A. & Gilbey, J. Genetic diversity and differentiation of Nile tilapia
3596 (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquaculture Research* **36**,
3597 1450-1457 (2005).

3598 44 Lin, G., Chua, E., Orban, L. & Yue, G. H. Mapping QTL for Sex and Growth Traits in Salt-
3599 Tolerant Tilapia (*Oreochromis spp.* X *O. mossambicus*). *PLOS One* **11**, e0166723 (2016).

3600 45 Xia, J. H. *et al.* Signatures of selection in tilapia revealed by whole genome resequencing.
3601 *Scientific Reports* **5**, 14168 (2015).

3602 46 Cnaani, A. *et al.* Detection of a chromosomal region with two quantitative trait loci,
3603 affecting cold tolerance and fish size, in an F2 tilapia hybrid. *Aquaculture* **223**, 117-128
3604 (2003).

3605 47 Joshi, R. *et al.* Development and validation of 58K SNP-array and high-density linkage
3606 map in Nile tilapia (*O. niloticus*). *Frontiers in Genetics* **9**, 472 (2018).

3607 48 Li, H. L. *et al.* Genome-wide QTL analysis identified significant associations between
3608 hypoxia tolerance and mutations in the GPR132 and ABCG4 genes in Nile tilapia.
3609 *Marine Biotechnology* **19**, 441-453 (2017).

3610 49 Brawand, D. *et al.* The genomic substrate for adaptive radiation in African cichlid fish.
3611 *Nature* **513**, 375 (2014).

3612 50 Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J. & Kocher, T. D. A high
3613 quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the
3614 structure of two sex determination regions. *BMC Genomics* **18**, 341 (2017).

3615 51 Bhatta, S. *et al.* Gonads directly regulate growth in teleosts. *Proceedings of the National*
3616 *Academy of Sciences* **109**, 11408-11412 (2012).

3617 52 Melamed, P. *et al.* Hypothalamic and thyroidal regulation of growth hormone in tilapia.
3618 *General and Comparative Endocrinology* **97**, 13-30 (1995).

3619 53 Fuentes, E. N. *et al.* The IGF-I/PI3K/Akt and the IGF-I/MAPK/ERK pathways in vivo in
3620 skeletal muscle are regulated by nutrition and contribute to somatic growth in the fine
3621 flounder. *American Journal of Physiology-Heart and Circulatory Physiology* (2011).

3622 54 Miura, C. *et al.* Gh is produced by the testis of Japanese eel and stimulates proliferation
3623 of spermatogonia. *Reproduction* **142**, 869-877 (2011).

3624 55 Pamela, N.-R., Maricela, L., Carlos, V.-R. & Aurea, O. 3,5-di-iodothyronine stimulates
3625 tilapia growth through an alternate isoform of thyroid hormone receptor $\beta 1$. *Journal*
3626 *of Molecular Endocrinology* **52**, 1 (2014).

3627 56 Ogino, Y. *et al.* Bmp7 and Lef1 are the downstream effectors of androgen signaling in
3628 androgen-induced sex characteristics development in medaka. *Endocrinology* **155**,
3629 449-462 (2014).

3630 57 Faught, E. & Vijayan, M. M. Mechanisms of cortisol action in fish hepatocytes.
3631 *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology*
3632 **199**, 136-145 (2016).

3633 58 Sharma, E. *et al.* Transcriptome assemblies for studying sex-biased gene expression in
3634 the guppy, *Poecilia reticulata*. *BMC Genomics* **15**, 400 (2014).

3635 59 Wong, R. Y., McLeod, M. M. & Godwin, J. Limited sex-biased neural gene expression
3636 patterns across strains in Zebrafish (*Danio rerio*). *BMC Genomics* **15**, 905 (2014).

3637 60 Beal, A. P., Martin, F. D. & Hale, M. C. Using RNA-seq to determine patterns of sex-bias
3638 in gene expression in the brain of the sex-role reversed Gulf Pipefish (*Syngnathus*
3639 *scovelli*). *Marine Genomics* **37**, 120-127 (2018).

3640 61 Shao, C. *et al.* Epigenetic modification and inheritance in sexual reversal of fish.
3641 *Genome Research* **24**, 604-615 (2014).

3642 62 Navarro-Martín, L. *et al.* DNA methylation of the gonadal aromatase (cyp19a) promoter
3643 is involved in temperature-dependent sex ratio shifts in the European sea bass. *PLOS*
3644 *Genetics* **7**, e1002447 (2011).

3645 63 Campos, C., Valente, L., Conceição, L., Engrola, S. & Fernandes, J. Temperature affects
3646 methylation of the myogenin putative promoter, its expression and muscle cellularity
3647 in Senegalese sole larvae. *Epigenetics* **8**, 389-397 (2013).

- Si, Y. *et al.* DNA methylation level of cyp19a1a and Foxl2 gene related to their expression patterns and reproduction traits during ovary development stages of Japanese flounder (*Paralichthys olivaceus*). *Gene* **575**, 321-330 (2016).
- Burgerhout, E. *et al.* Genetic background and embryonic temperature affect DNA methylation and expression of myogenin and muscle development in Atlantic salmon (*Salmo salar*). *PLOS One* **12**, e0179918 (2017).
- Pollux, B. J. A., Meredith, R. W., Springer, M. S., Garland, T. & Reznick, D. N. The evolution of the placenta drives a shift in sexual selection in livebearing fish. *Nature* **513**, 233 (2014).
- Maan, M. E. & Taborsky, M. Sexual conflict over breeding substrate causes female expulsion and offspring loss in a cichlid fish. *Behavioral Ecology* **19**, 302-308 (2007).
- van Doorn, G. S. & Kirkpatrick, M. Turnover of sex chromosomes induced by sexual conflict. *Nature* **449**, 909 (2007).
- Ritchie, M. G., Webb, S. A., Graves, J. A., Magurran, A. E. & Macias, G. C. Patterns of speciation in endemic Mexican Goodeid fish: sexual conflict or early radiation? *Journal of Evolutionary Biology* **18**, 922-929 (2005).
- Magurran, A. E. & Seghers, B. H. Sexual conflict as a consequence of ecology: evidence from guppy, *Poecilia reticulata*, populations in Trinidad. *Proceedings of the Royal Society of London. Series B: Biological Sciences* **255**, 31-36 (1994).
- Kitahashi, T., Shahjahan, M. & Parhar, I. S. in *Sexual Plasticity and Gametogenesis in Fishes* (ed Balasubramanian Senthilkumaran) Ch. Hypothalamic Regulation of Pituitary Gonadotropins, (Nova Science Publishers, 2013).
- Canosa, L. F., Chang, J. P. & Peter, R. E. Neuroendocrine control of growth hormone in fish. *General and Comparative Endocrinology* **151**, 1-26 (2007).
- Reinecke, M. *et al.* Growth hormone and insulin-like growth factors in fish: where we are and where to go. *General and Comparative Endocrinology* **142**, 20-24 (2005).
- Holloway, A. C. & Leatherland, J. F. Neuroendocrine Regulation of Growth Hormone Secretion in Teleost Fishes with Emphasis on the Involvement of Gonadal Sex Steroids. *Reviews in Fish Biology and Fisheries* **8**, 409-429 (1998).
- Davis, L. K. *et al.* Gender-specific expression of multiple estrogen receptors, growth hormone receptors, insulin-like growth factors and vitellogenins, and effects of 17 β -estradiol in the male tilapia (*Oreochromis mossambicus*). *General and Comparative Endocrinology* **156**, 544-551 (2008).
- Sakamoto, T. *et al.* Osmoregulatory actions of growth hormone and prolactin in an advanced teleost. *General and Comparative Endocrinology* **106**, 95-101 (1997).
- Mancera, J. M. & McCormick, S. D. Evidence for growth hormone/insulin-like growth factor I axis regulation of seawater acclimation in the euryhaline teleost *Fundulus heteroclitus*. *General and Comparative Endocrinology* **111**, 103-112 (1998).
- Shved, N., Berishvili, G., Mazel, P., Baroiller, J. F. & Eppler, E. Growth hormone (GH) treatment acts on the endocrine and autocrine/paracrine GH/IGF-axis and on TNF- α expression in bony fish pituitary and immune organs. *Fish & Shellfish Immunology* **31**, 944-952 (2011).
- Perez-Sanchez, J. The involvement of growth hormone in growth regulation, energy homeostasis and immune function in the gilthead sea bream (*Sparus aurata*): a short review. *Fish Physiology and Biochemistry* **22**, 135-144 (2000).
- Canosa, L. F., Unniappan, S. & Peter, R. E. Periprandial changes in growth hormone release in goldfish: Role of somatostatin, ghrelin and gastrin releasing peptide. *American Journal of Physiology-Regulatory, Integrative and Comparative Physiology* (2005).
- Johnston, I. A., Bower, N. I. & Macqueen, D. J. Growth and the regulation of myotomal muscle mass in teleost fish. *Journal of Experimental Biology* **214**, 1617-1628 (2011).
- Hayward, A. & Gillooly, J. F. The cost of sex: quantifying energetic investment in gamete production by males and females. *PLOS One* **6**, e16557 (2011).

3701 83 Lohr, H. *et al.* Diet-Induced Growth Is Regulated via Acquired Leptin Resistance and
3702 Engages a Pomc-Somatostatin-Growth Hormone Circuit. *Cell reports* **23**, 1728-1741
3703 (2018).
3704 84 Uchida, K. *et al.* Effects of fasting on growth hormone/insulin-like growth factor I axis
3705 in the tilapia, *Oreochromis mossambicus*. *Comparative Biochemistry and Physiology*
3706 *Part A: Molecular & Integrative Physiology* **134**, 429-439 (2003).
3707 85 Polakof, S., Panserat, S., Soengas, J. L. & Moon, T. W. Glucose metabolism in fish: a
3708 review. *Journal of Comparative Physiology B* **182**, 1015-1045 (2012).
3709 86 Reindl, K. M. & Sheridan, M. A. Peripheral regulation of the growth hormone-insulin-
3710 like growth factor system in fish and other vertebrates. *Comparative Biochemistry and*
3711 *Physiology Part A: Molecular & Integrative Physiology* **163**, 231-245 (2012).
3712 87 Jason, S. L. & Cui, W. Proliferation, survival and metabolism: the role of PI3K/AKT/mTOR
3713 signalling in pluripotency and cell fate determination. *Development (Cambridge,*
3714 *England)* **143**, 3050-3060 (2016).
3715 88 Saxton, R. A. & Sabatini, D. M. mTOR signaling in growth, metabolism, and disease. *Cell*
3716 **168**, 960-976 (2017).
3717 89 Zhang, W. & Liu, H. T. MAPK signal pathways in the regulation of cell proliferation in
3718 mammalian cells. *Cell Research* **12**, 9 (2002).
3719 90 Pozios, K. C., Ding, J., Degger, B., Upton, Z. & Duan, C. IGFs stimulate zebrafish cell
3720 proliferation by activating MAP kinase and PI3-kinase-signaling pathways. *American*
3721 *Journal of Physiology-Regulatory, Integrative and Comparative Physiology* **280**, R1230-
3722 R1239 (2001).
3723 91 Grabherr, M. G. *et al.* Trinity: reconstructing a full-length transcriptome without a
3724 genome from RNA-Seq data. *Nature Biotechnology* **29**, 644 (2011).
3725 92 Penny, G. D., Kay, G. F., Sheardown, S. A., Rastan, S. & Brockdorff, N. Requirement for
3726 Xist in X chromosome inactivation. *Nature* **379**, 131 (1996).
3727 93 Mank, J. E. & Ellegren, H. All dosage compensation is local: gene-by-gene regulation of
3728 sex-biased expression on the chicken Z chromosome. *Heredity* **102**, 312 (2009).
3729 94 Meiklejohn, C. D., Coolon, J. D., Hartl, D. L. & Wittkopp, P. J. The roles of cis- and trans-
3730 regulation in the evolution of regulatory incompatibilities and sexually dimorphic gene
3731 expression. *Genome Research* **24**, 84-95 (2014).
3732 95 Stocks, M., Dean, R., Rogell, B. & Friberg, U. Sex-specific trans-regulatory variation on
3733 the *Drosophila melanogaster* X chromosome. *PLOS Genetics* **11**, e1005015 (2015).
3734 96 Harrison, P. W. *et al.* Sexual selection drives evolution and rapid turnover of male gene
3735 expression. *Proceedings of the National Academy of Sciences* **112**, 4393-4398 (2015).
3736 97 Mank, J. E., Hultin-Rosenberg, L., Axelsson, E. & Ellegren, H. Rapid evolution of female-
3737 biased, but not male-biased, genes expressed in the avian brain. *Molecular Biology and*
3738 *Evolution* **24**, 2698-2706 (2007).
3739 98 Yang, L. D., Zhang, Z. L. & He, S. P. Both male-biased and female-biased genes evolve
3740 faster in fish genomes. *Genome Biology and Evolution* **8**, 3433-3445 (2016).
3741 99 Swanson, W. J., Wong, A., Wolfner, M. F. & Aquadro, C. F. Evolutionary expressed
3742 sequence tag analysis of *Drosophila* female reproductive tracts identifies genes
3743 subjected to positive selection. *Genetics* **168**, 1457-1465 (2004).
3744 100 Wittkopp, P. J., Haerum, B. K. & Clark, A. G. Evolutionary changes in cis and trans gene
3745 regulation. *Nature* **430**, 85 (2004).
3746 101 Bird, A. The essentials of DNA methylation. *Cell* **70**, 5-8 (1992).
3747 102 Suzuki, M. M. & Bird, A. DNA methylation landscapes: provocative insights from
3748 epigenomics. *Nature Reviews Genetics* **9**, 465 (2008).
3749 103 Burger, L., Gaidatzis, D., Schubeler, D. & Stadler, M. B. Identification of active regulatory
3750 regions from DNA methylation data. *Nucleic Acids Research* **41**, e155 (2013).
3751 104 Hon, G. C. *et al.* Epigenetic memory at embryonic enhancers identified in DNA
3752 methylation maps from adult mouse tissues. *Nature Genetics* **45**, 1198 (2013).
3753 105 Schübeler, D. Function and information content of DNA methylation. *Nature* **517**, 321
3754 (2015).

3755 106 Akalin, A. *et al.* methylKit: a comprehensive R package for the analysis of genome-wide
3756 DNA methylation profiles. *Genome Biology* **13**, R87-R87 (2012).
3757 107 Mank, J. E. Population genetics of sexual conflict in the genomic era. *Nature Reviews*
3758 *Genetics* **18**, 721 (2017).
3759 108 Mank, J. E. The transcriptional architecture of phenotypic dimorphism. *Nature Ecology*
3760 *& Evolution* **1**, 0006 (2017).
3761 109 Arnqvist, G. & Rowe, L. *Sexual Conflict*. (Princeton University Press, 2005).
3762 110 Stewart, A. D., Morrow, E. H. & Rice, W. R. Assessing putative interlocus sexual conflict
3763 in *Drosophila melanogaster* using experimental evolution. *Proceedings of the Royal*
3764 *Society of London. Series B: Biological Sciences* **272**, 2029-2035 (2005).
3765 111 Zhang, Z., Hambuch, T. M. & Parsch, J. Molecular evolution of sex-biased genes in
3766 *Drosophila*. *Molecular Biology and Evolution* **21**, 2130-2139 (2004).
3767 112 Ellegren, H. & Parsch, J. The evolution of sex-biased genes and sex-biased gene
3768 expression. *Nature Reviews Genetics* **8**, 689-698 (2007).
3769 113 Baines, J. F., Sawyer, S. A., Hartl, D. L. & Parsch, J. Effects of X-linkage and sex-biased
3770 gene expression on the rate of adaptive protein evolution in *Drosophila*. *Molecular*
3771 *Biology and Evolution* **25**, 1639-1650 (2008).
3772 114 Connallon, T. & Clark, A. G. Association between sex-biased gene expression and
3773 mutations with sex-specific phenotypic consequences in *Drosophila*. *Genome Biology*
3774 *and Evolution* **3**, 151-155 (2011).
3775 115 Parsch, J. & Ellegren, H. The evolutionary causes and consequences of sex-biased gene
3776 expression. *Nature Reviews Genetics* **14**, 83-87 (2013).
3777 116 Perry, J. C., Harrison, P. W. & Mank, J. E. The ontogeny and evolution of sex-biased gene
3778 expression in *Drosophila melanogaster*. *Molecular Biology and Evolution* **31**, 1206-1219
3779 (2014).
3780 117 Rice, W. R. & Holland, B. The enemies within: intergenomic conflict, interlocus contest
3781 evolution (ICE), and the intraspecific Red Queen. *Behavioral Ecology and Sociobiology*
3782 **41**, 1-10 (1997).
3783 118 Metzker, M. L. Sequencing technologies — the next generation. *Nature Reviews*
3784 *Genetics* **11**, 31 (2009).
3785 119 Meissner, A. *et al.* Reduced representation bisulfite sequencing for comparative high-
3786 resolution DNA methylation analysis. *Nucleic Acids Research* **33**, 5868-5877 (2005).
3787 120 Ozsolak, F. & Milos, P. M. RNA sequencing: advances, challenges and opportunities.
3788 *Nature Reviews Genetics* **12**, 87 (2010).
3789 121 Laird, P. W. Principles and challenges of genome-wide DNA methylation analysis.
3790 *Nature Reviews Genetics* **11**, 191 (2010).
3791 122 Frommer, M. *et al.* A genomic sequencing protocol that yields a positive display of 5-
3792 methylcytosine residues in individual DNA strands. *Proceedings of the National*
3793 *Academy of Sciences* **89**, 1827 (1992).
3794 123 Susan, J. C., Harrison, J., Paul, C. L. & Frommer, M. High sensitivity mapping of
3795 methylated cytosines. *Nucleic Acids Research* **22**, 2990-2997 (1994).
3796 124 Robertson, M. & Richards, C. Opportunities and challenges of next-generation
3797 sequencing applications in ecological epigenetics. *Molecular Ecology* **24**, 3799-3801
3798 (2015).
3799 125 Wang, Z., Gerstein, M. & Snyder, M. RNA-Seq: a revolutionary tool for transcriptomics.
3800 *Nature Reviews Genetics* **10**, 57 (2009).
3801 126 Martin, J. A. & Wang, Z. Next-generation transcriptome assembly. *Nature Reviews*
3802 *Genetics* **12**, 671 (2011).
3803 127 Byron, S. A., Van Keuren-Jensen, K. R., Engelthaler, D. M., Carpten, J. D. & Craig, D. W.
3804 Translating RNA sequencing into clinical diagnostics: opportunities and challenges.
3805 *Nature Reviews Genetics* **17**, 257 (2016).
3806 128 Berger, B., Peng, J. & Singh, M. Computational solutions for omics data. *Nature Reviews*
3807 *Genetics* **14**, 333 (2013).
3808 129 Kim, D. *et al.* TopHat2: accurate alignment of transcriptomes in the presence of
3809 insertions, deletions and gene fusions. *Genome Biology* **14**, R36 (2013).

3810 130 Dobin, A. *et al.* STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**, 15-21
3811 (2013).
3812 131 Grabherr, M. G. *et al.* Full-length transcriptome assembly from RNA-Seq data without
3813 a reference genome. *Nature Biotechnology* **29**, 644 (2011).
3814 132 Love, M. I., Huber, W. & Anders, S. Moderated estimation of fold change and dispersion
3815 for RNA-seq data with DESeq2. *Genome Biology* **15**, 550 (2014).
3816 133 McCarthy, D. J., Chen, Y. & Smyth, G. K. Differential expression analysis of multifactor
3817 RNA-Seq experiments with respect to biological variation. *Nucleic Acids Research* **40**,
3818 4288-4297 (2012).
3819 134 Trapnell, C. *et al.* Differential gene and transcript expression analysis of RNA-seq
3820 experiments with TopHat and Cufflinks. *Nature Protocols* **7**, 562-578 (2012).
3821 135 Bard, J. B. L. & Rhee, S. Y. Ontologies in biology: design, applications and future
3822 challenges. *Nature Reviews Genetics* **5**, 213-222 (2004).
3823 136 Yang, Z. PAML 4: Phylogenetic Analysis by Maximum Likelihood. *Molecular Biology and*
3824 *Evolution* **24**, 1586-1591 (2007).
3825 137 Yang, Z. Likelihood ratio tests for detecting positive selection and application to
3826 primate lysozyme evolution. *Molecular biology and evolution* **15**, 568-573 (1998).
3827 138 Kryazhimskiy, S. & Plotkin, J. B. The population genetics of dN/dS. *PLOS Genetics* **4**,
3828 e1000304 (2008).
3829 139 Zhang, C., Wang, J., Long, M. & Fan, C. gKaKs: the pipeline for genome-level Ka/Ks
3830 calculation. *Bioinformatics* **29**, 645-646 (2013).
3831 140 Bustamante, C. D. *et al.* Natural selection on protein-coding genes in the human
3832 genome. *Nature* **437**, 1153-1157 (2005).
3833 141 Yang, Z. & Bielawski, J. P. Statistical methods for detecting molecular adaptation.
3834 *Trends in Ecology & Evolution* **15**, 496-503 (2000).
3835 142 Murrell, B. *et al.* Gene-Wide identification of episodic selection. *Molecular Biology and*
3836 *Evolution* **32**, 1365-1371 (2015).
3837 143 Kim, H. & Kim, J.-S. A guide to genome engineering with programmable nucleases.
3838 *Nature Reviews Genetics* **15**, 321 (2014).
3839 144 Housden, B. E. *et al.* Loss-of-function genetic tools for animal models: cross-species and
3840 cross-platform differences. *Nature Reviews Genetics* **18**, 24 (2016).
3841 145 Kim, Y. G., Cha, J. & Chandrasegaran, S. Hybrid restriction enzymes: zinc finger fusions
3842 to Fok I cleavage domain. *Proceedings of the National Academy of Sciences* **93**, 1156
3843 (1996).
3844 146 Kim, Y. *et al.* A library of TAL effector nucleases spanning the human genome. *Nature*
3845 *Biotechnology* **31**, 251 (2013).
3846 147 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*
3847 **8**, 2281-2308 (2013).
3848 148 Li, M. *et al.* Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics*
3849 **197**, 591-599 (2014).
3850 149 Hwang, W. Y. *et al.* Efficient genome editing in zebrafish using a CRISPR-Cas system.
3851 *Nature Biotechnology* **31**, 227 (2013).
3852 150 Yin, L., Jao, L. E. & Chen, W. Generation of Targeted Mutations in Zebrafish Using the
3853 CRISPR/Cas System. *Methods in Molecular Biology* **1332**, 205-217 (2015).
3854 151 Doudna, J. A. & Charpentier, E. The new frontier of genome engineering with CRISPR-
3855 Cas9. *Science* **346**, 1258096 (2014).
3856 152 Ansai, S. & Kinoshita, M. Targeted mutagenesis using CRISPR/Cas system in medaka.
3857 *Biology Open* **3**, 362-371 (2014).
3858 153 Juntti, S. A. *et al.* A neural basis for control of cichlid female reproductive behavior by
3859 prostaglandin F2 α . *Current Biology* **26**, 943-949 (2016).
3860 154 Edvardsen, R. B., Leininger, S., Kleppe, L., Skaftnesmo, K. O. & Wargelius, A. Targeted
3861 mutagenesis in Atlantic salmon (*Salmo salar* L.) using the CRISPR/Cas9 system induces
3862 complete knockout individuals in the F0 generation. *PLOS One* **9**, e108622 (2014).
3863 155 Li, M. *et al.* Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics*
3864 **197**, 591-599 (2014).

3865 156 Kishimoto, K. *et al.* Production of a breed of red sea bream *Pagrus major* with an
3866 increase of skeletal muscle mass and reduced body length by genome editing with
3867 CRISPR/Cas9. *Aquaculture* **495**, 415-427 (2018).

3868 157 Khalil, K. *et al.* Generation of myostatin gene-edited channel catfish (*Ictalurus*
3869 *punctatus*) via zygote injection of CRISPR/Cas9 system. *Scientific Reports* **7**, 7301 (2017).

3870 158 Bae, S., Park, J. & Kim, J. S. Cas-OFFinder: a fast and versatile algorithm that searches
3871 for potential off-target sites of Cas9 RNA-guided endonucleases. *Bioinformatics* **30**,
3872 1473-1475 (2014).

3873 159 Park, J., Bae, S. & Kim, J.-S. Cas-Designer: a web-based tool for choice of CRISPR-Cas9
3874 target sites. *Bioinformatics* **31**, 4014-4016 (2015).

3875 160 Dehler, C. E., Boudinot, P., Martin, S. A. M. & Collet, B. Development of an Efficient
3876 Genome Editing Method by CRISPR/Cas9 in a Fish Cell Line. *Marine Biotechnology* **18**,
3877 449-452 (2016).

3878 161 Liu, Q., Yuan, Y., Zhu, F., Hong, Y. & Ge, R. Efficient genome editing using CRISPR/Cas9
3879 ribonucleoprotein approach in cultured Medaka fish cells. *Biology Open* **7**, bio035170
3880 (2018).

3881 162 Diotel, N. *et al.* Aromatase in the brain of teleost fish: expression, regulation and
3882 putative functions. *Frontiers in Neuroendocrinology* **31**, 172-192 (2010).

3883 163 Karigo, T. & Oka, Y. Neurobiological study of fish brains gives insights into the nature of
3884 gonadotropin-releasing hormone 1-3 neurons. *Front Endocrinol (Lausanne)* **4**, 177-177
3885 (2013).

3886 164 Harris, R. M., Dijkstra, P. D. & Hofmann, H. A. Complex structural and regulatory
3887 evolution of the pro-opiomelanocortin gene family. *General and Comparative*
3888 *Endocrinology* **195**, 107-115 (2014).

3889 165 Spicer, O. S., Wong, T.-T., Zmora, N. & Zohar, Y. Targeted Mutagenesis of the
3890 Hypophysiotropic Gnrh3 in Zebrafish (*Danio rerio*) Reveals No Effects on Reproductive
3891 Performance. *PLOS One* **11**, e0158141 (2016).

3892 166 Biebertmann, H. *et al.* A role for β -melanocyte-stimulating hormone in human body-
3893 weight regulation. *Cell Metabolism* **3**, 141-146 (2006).

3894 167 Corander, M. P., Rimmington, D., Challis, B. G., O'Rahilly, S. & Coll, A. P. Loss of agouti-
3895 related peptide does not significantly impact the phenotype of murine POMC
3896 deficiency. *Endocrinology* **152**, 1819-1828 (2011).

3897 168 Erica, J. P. A. *et al.* 60 YEARS OF POMC: Regulation of feeding and energy homeostasis
3898 by α -MSH. *Journal of Molecular Endocrinology* **56**, T157-T174 (2016).

3899 169 Gallo-Payet, N. 60 years of POMC: adrenal and extra-adrenal functions of ACTH. *Journal*
3900 *of Molecular Endocrinology* **56**, T135-T156 (2016).

3901 170 Millington, G. W. The role of proopiomelanocortin (POMC) neurones in feeding
3902 behaviour. *Nutr Metab (Lond)* **4**, 18-18 (2007).

3903 171 Murashita, K. *et al.* Leptin reduces Atlantic salmon growth through the central pro-
3904 opiomelanocortin pathway. *Comparative Biochemistry and Physiology Part A:*
3905 *Molecular & Integrative Physiology* **158**, 79-86 (2011).

3906 172 Wang, C. *et al.* Tap63 contributes to sexual dimorphism in POMC neuron functions and
3907 energy homeostasis. *Nature Communications* **9**, 1544 (2018).

3908 173 Fryer, J. N., Nishioka, R. S. & Bern, H. A. Somatostatin inhibition of teleost growth
3909 hormone secretion. *General and Comparative Endocrinology* **39**, 244-246 (1979).

3910 174 Ocampo Daza, D., Sundström, G., Bergqvist, C. A. & Larhammar, D. The evolution of
3911 vertebrate somatostatin receptors and their gene regions involves extensive
3912 chromosomal rearrangements. *BMC Evolutionary Biology* **12**, 231-231 (2012).

3913 175 Xu, J. *et al.* Genome-wide SNP discovery from transcriptome of four common carp
3914 strains. *PLOS One* **7**, e48140 (2012).

3915 176 Popma, T. & Masser, M. Tilapia: Life History and Biology. (Stoneville, Mississippi:
3916 Southern Regional Aquaculture Center, 1999).

3917 177 Yue, G. H. & Orban, L. A simple and affordable method for high-throughput DNA
3918 extraction from animal tissues for polymerase chain reaction. *Electrophoresis* **26**, 3081-
3919 3083 (2005).

3920 178 Smit, A. F. A., Hubley, R. & Green, P. *RepeatMasker Open-4.0*. (2015).
3921 179 Bailey, T. L. *et al.* MEME SUITE: tools for motif discovery and searching. *Nucleic Acids*
3922 *Research* **37**, W202-W208 (2009).
3923 180 Carleton, K. L. *et al.* Chromosome-scale assemblies reveal the structural evolution of
3924 African cichlid genomes. *GigaScience* **8** (2019).
3925 181 Bolger, A. M., Lohse, M. & Usadel, B. Trimmomatic: a flexible trimmer for Illumina
3926 sequence data. *Bioinformatics* **30**, 2114-2120 (2014).
3927 182 Anders, S., Pyl, P. T. & Huber, W. HTSeq--a Python framework to work with high-
3928 throughput sequencing data. *Bioinformatics* **31**, 166-169 (2015).
3929 183 Yu, G., Wang, L. G., Han, Y. & He, Q. Y. clusterProfiler: an R package for comparing
3930 biological themes among gene clusters. *Omics : a journal of integrative biology* **16**, 284-
3931 287 (2012).
3932 184 Mi, H., Muruganujan, A., Ebert, D., Huang, X. & Thomas, P. D. PANTHER version 14:
3933 more genomes, a new PANTHER GO-slim and improvements in enrichment analysis
3934 tools. *Nucleic Acids Research* **47**, D419-d426 (2019).
3935 185 McKenna, A. *et al.* The Genome Analysis Toolkit: a MapReduce framework for analyzing
3936 next-generation DNA sequencing data. *Genome Research* **20**, 1297-1303 (2010).
3937 186 Li, H. A statistical framework for SNP calling, mutation discovery, association mapping
3938 and population genetical parameter estimation from sequencing data. *Bioinformatics*
3939 **27**, 2987-2993 (2011).
3940 187 Li, H. *et al.* The Sequence Alignment/Map format and SAMtools. *Bioinformatics* **25**,
3941 2078-2079 (2009).
3942 188 Zerbino, D. R. *et al.* Ensembl 2018. *Nucleic Acids Research* **46**, D754-D761 (2018).
3943 189 Rice, P., Longden, I. & Bleasby, A. EMBOSS: the European Molecular Biology Open
3944 Software Suite. *Trends in Genetics* **16**, 276-277 (2000).
3945 190 Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment
3946 search tool. *Journal of molecular biology* **215**, 403-410 (1990).
3947 191 Katoh, K. & Standley, D. M. MAFFT multiple sequence alignment software version 7:
3948 improvements in performance and usability. *Molecular Biology and Evolution* **30**, 772-
3949 780 (2013).
3950 192 Suyama, M., Torrents, D. & Bork, P. PAL2NAL: robust conversion of protein sequence
3951 alignments into the corresponding codon alignments. *Nucleic Acids Research* **34**,
3952 W609-W612 (2006).
3953 193 Gardell, A. M., Qin, Q., Rice, R. H., Li, J. & Kültz, D. Derivation and osmotolerance
3954 characterization of three immortalized tilapia (*Oreochromis mossambicus*) cell lines.
3955 *PLOS One* **9**, e95919-e95919 (2014).
3956 194 Ran, F. A. *et al.* Genome engineering using the CRISPR-Cas9 system. *Nature Protocols*
3957 **8**, 2281 (2013).
3958 195 Kumar, S., Stecher, G., Li, M., Knyaz, C. & Tamura, K. MEGA X: molecular evolutionary
3959 genetics analysis across computing platforms. *Molecular Biology and Evolution* **35**,
3960 1547-1549 (2018).
3961 196 Ho, J., Tumkaya, T., Aryal, S., Choi, H. & Claridge-Chang, A. in *bioRxiv* 377978 (2019).
3962 197 Patel, R. K. & Jain, M. NGS QC Toolkit: A Toolkit for Quality Control of Next Generation
3963 Sequencing Data. *PLOS One* **7**, e30619 (2012).
3964 198 Krueger, F. & Andrews, S. R. Bismark: a flexible aligner and methylation caller for
3965 Bisulfite-Seq applications. *Bioinformatics* **27**, 1571-1572 (2011).
3966 199 Langmead, B. & Salzberg, S. L. Fast gapped-read alignment with Bowtie 2. *Nature*
3967 *Methods* **9**, 357-359 (2012).
3968 200 Jones, P. A. & Takai, D. The Role of DNA Methylation in Mammalian Epigenetics. *Science*
3969 **293**, 1068 (2001).
3970 201 Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and
3971 animals. *Proceedings of the National Academy of Sciences* **107**, 8689-8694 (2010).
3972 202 Zemach, A., McDaniel, I. E., Silva, P. & Zilberman, D. Genome-Wide Evolutionary
3973 Analysis of Eukaryotic DNA Methylation. *Science* **328**, 916-919 (2010).

3974 203 Mathelier, A. *et al.* JASPAR 2014: an extensively expanded and updated open-access
3975 database of transcription factor binding profiles. *Nucleic Acids Research*, gkt997 (2013).
3976 204 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal
3977 regulatory regions. *Nature* **480**, 490 (2011).
3978 205 Raveh, S., Gavert, N., Spiegel, I. & Ben-Ze'ev, A. The cell adhesion nectin-like molecules
3979 (Nect) 1 and 4 suppress the growth and tumorigenic ability of colon cancer cells. *Journal*
3980 *of Cellular Biochemistry* **108**, 326-336 (2009).
3981 206 Mi, H. *et al.* PANTHER version 7: improved phylogenetic trees, orthologs and
3982 collaboration with the Gene Ontology Consortium. *Nucleic Acids Research* **38**, D204-
3983 D210 (2010).
3984 207 Nydam, M. L. & Harrison, R. G. Reproductive protein evolution in two cryptic species of
3985 marine chordate. *BMC Evolutionary Biology* **11**, 18 (2011).
3986 208 Lipinska, A. *et al.* Sexual dimorphism and the evolution of sex-biased gene expression
3987 in the brown alga *Ectocarpus*. *Molecular Biology and Evolution* **32**, 1581-1597 (2015).
3988 209 Tao, W. J. *et al.* Characterization of Gonadal Transcriptomes from Nile Tilapia
3989 (*Oreochromis niloticus*) Reveals Differentially Expressed Genes. *PLOS One* **8**, e63604
3990 (2013).
3991 210 Tao, W. J. *et al.* Transcriptome display during tilapia sex determination and
3992 differentiation as revealed by RNA-Seq analysis. *BMC Genomics* **19**, 363 (2018).
3993 211 Neves, S. R., Ram, P. T. & Iyengar, R. G Protein Pathways. *Science* **296**, 1636 (2002).
3994 212 Gore, A. C. in *Fundamental Neuroscience (Fourth Edition)* (eds Larry R. Squire *et al.*)
3995 799-817 (2013).
3996 213 Jenks, B. G. Regulation of Proopiomelanocortin Gene Expression. *Annals of the New*
3997 *York Academy of Sciences* **1163**, 17-30 (2009).
3998 214 Williams, L. M., Ma, X., Boyko, A. R., Bustamante, C. D. & Oleksiak, M. F. SNP
3999 identification, verification, and utility for population genetics in a non-model genus.
4000 *BMC Genetics* **11**, 32 (2010).
4001 215 Reue, K. & Donkor, J. Lipin: a determinant of adiposity, insulin sensitivity and energy
4002 balance. *Future Lipidology* **1**, 91-101 (2006).
4003 216 Borrás, C. *et al.* RasGrf1 deficiency delays aging in mice. *Aging* **3**, 262-276 (2011).
4004 217 Raffan, E. *et al.* A Deletion in the canine POMC gene is associated with weight and
4005 appetite in obesity-prone labrador retriever dogs. *Cell Metabolism* **23**, 893-900 (2016).
4006 218 Yaswen, L., Diehl, N., Brennan, M. B. & Hochgeschwender, U. Obesity in the mouse
4007 model of pro-opiomelanocortin deficiency responds to peripheral melanocortin.
4008 *Nature Medicine* **5**, 1066 (1999).
4009 219 Schulte, P. M. & Metzger, D. C. H. The DNA methylation landscape of stickleback reveals
4010 patterns of sex chromosome evolution and effects of environmental salinity. *Genome*
4011 *Biology and Evolution* **10**, 775-785 (2018).
4012 220 Ziller, M. J. *et al.* Charting a dynamic DNA methylation landscape of the human genome.
4013 *Nature* **500**, 477 (2013).
4014 221 Lister, R. *et al.* Global epigenomic reconfiguration during mammalian brain
4015 development. *Science* **341**, 1237905 (2013).
4016 222 Hodges, E. *et al.* Directional DNA methylation changes and complex intermediate states
4017 accompany lineage specificity in the adult hematopoietic compartment. *Molecular Cell*
4018 **44**, 17-28 (2011).
4019 223 Feng, S. *et al.* Conservation and divergence of methylation patterning in plants and
4020 animals. *P. Natl. Acad. Sci. USA* **107**, 8689-8694 (2010).
4021 224 Jeltsch, A. Phylogeny of methylomes. *Science* **328**, 837-838 (2010).
4022 225 Deniz, Ö., Frost, J. M. & Branco, M. R. Regulation of transposable elements by DNA
4023 modifications. *Nature Reviews Genetics* **20**, 417-431 (2019).
4024 226 Chernyavskaya, Y. *et al.* Loss of DNA methylation in zebrafish embryos activates
4025 retrotransposons to trigger antiviral signaling. *Development (Cambridge, England)* **144**,
4026 2925-2939 (2017).
4027 227 Stadler, M. B. *et al.* DNA-binding factors shape the mouse methylome at distal
4028 regulatory regions. *Nature* **480**, 490-495 (2011).

4029 228 Burger, L., Gaidatzis, D., Schuebeler, D. & Stadler, M. B. Identification of active
 4030 regulatory regions from DNA methylation data. *Nucleic Acids Research* **41**, e155 (2013).
 4031 229 Jones, P. A. Functions of DNA methylation: islands, start sites, gene bodies and beyond.
 4032 *Nature Reviews Genetics* **13**, 484-492 (2012).
 4033 230 Smith, S., Bernatchez, L. & Beheregaray, L. B. RNA-seq analysis reveals extensive
 4034 transcriptional plasticity to temperature stress in a freshwater fish species. *BMC*
 4035 *Genomics* **14**, 375 (2013).
 4036 231 Li, H. L., Lin, H. R. & Xia, J. H. Differential Gene Expression Profiles and Alternative
 4037 Isoform Regulations in Gill of Nile Tilapia in Response to Acute Hypoxia. *Marine*
 4038 *Biotechnology* **19**, 551-562 (2017).
 4039 232 Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J. & Kocher, T. D. A high
 4040 quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the
 4041 structure of two sex determination regions. *BMC Genomics* **18**, 341 (2017).
 4042 233 Font de Mora, J. *et al.* Ras-GRF1 signaling is required for normal beta-cell development
 4043 and glucose homeostasis. *EMBO Journal* **22**, 3039-3049 (2003).
 4044 234 Shi, X. Z. *et al.* Foxk1 promotes cell proliferation and represses myogenic differentiation
 4045 by regulating Foxo4 and Mef2. *Journal of Cell Science* **125**, 5329-5337 (2012).
 4046 235 Lin, X. *et al.* Dysregulation of insulin receptor substrate 2 in beta cells and brain causes
 4047 obesity and diabetes. *The Journal of Clinical Investigation* **114**, 908-916 (2004).
 4048 236 Tang, W. J., Fernandez, J. G., Sohn, J. J. & Amemiya, C. T. Chitin is endogenously
 4049 produced in vertebrates. *Current Biology* **25**, 897-900 (2015).
 4050 237 Reinius, B. *et al.* Female-biased expression of long non-coding RNAs in domains that
 4051 escape X-inactivation in mouse. *BMC Genomics* **11**, 614 (2010).
 4052 238 Cech, T. R. & Steitz, J. A. The noncoding RNA revolution - Trashing old rules to forge
 4053 new ones. *Cell* **157**, 77-94 (2014).
 4054 239 Collins, L. J., Schönfeld, B. & Chen, X. S. in *Handbook of Epigenetics* (ed Trygve
 4055 Tollefsbol) 49-61 (2011).
 4056 240 Arrigoni, A. *et al.* Analysis RNA-seq and Noncoding RNA. *Methods in Molecular Biology*
 4057 **1480**, 125-135 (2016).
 4058 241 Davis, P. J., Goglia, F. & Leonard, J. L. Nongenomic actions of thyroid hormone. *Nature*
 4059 *Reviews Endocrinology* **12**, 111 (2016).
 4060 242 Parhar, I. S., Soga, T. & Sakuma, Y. In situ hybridization for two differentially expressed
 4061 GnRH genes following estrogen and triiodothyronine treatment in the brains of juvenile
 4062 tilapia (cichlid). *Neuroscience Letters* **218**, 135-138 (1996).
 4063 243 Matsuyama, M., Selvaraj, S., Nyuji, M. & Ohga, H. in *Sexual Plasticity and*
 4064 *Gametogenesis in Fishes* (2013).
 4065 244 Kobayashi, Y., Nagahama, Y. & Nakamura, M. Diversity and plasticity of sex
 4066 determination and differentiation in fishes. *Sexual Development* **7**, 115-125 (2013).
 4067 245 Oka, Y. Three types of gonadotrophin-releasing hormone neurones and steroid -
 4068 sensitive sexually dimorphic kisspeptin neurones in teleosts. *Journal of*
 4069 *Neuroendocrinology* **21**, 334-338 (2009).
 4070 246 Parhar, I. S., Soga, T., Ogawa, S. & Sakuma, Y. FSH and LH- β subunits in the preoptic
 4071 nucleus: ontogenic expression in teleost. *General and Comparative Endocrinology* **132**,
 4072 369-378 (2003).
 4073 247 Ogawa, S. *et al.* Immunoneutralization of gonadotropin-releasing hormone type-III
 4074 suppresses male reproductive behavior of cichlids. *Neuroscience Letters* **403**, 201-205
 4075 (2006).
 4076 248 Chen, C.-C. & Fernald, R. D. Visual Information Alone Changes Behavior and Physiology
 4077 during Social Interactions in a Cichlid Fish (*Astatotilapia burtoni*). *PLOS One* **6**, e20313
 4078 (2011).
 4079 249 Janzen, W. J., Duncan, C. A. & Riley, L. G. Cortisol treatment reduces ghrelin signaling
 4080 and food intake in tilapia, *Oreochromis mossambicus*. *Domestic Animal Endocrinology*
 4081 **43**, 251-259 (2012).

4082 250 Andrew, L. P., Jason, P. B., Shunsuke, M., Tetsuya, H. & Grau, E. G. Differential
4083 regulation of Igf1 and Igf2 mRNA levels in tilapia hepatocytes: effects of insulin and
4084 cortisol on GH sensitivity. *Journal of Endocrinology* **211**, 201-210 (2011).
4085 251 Wagle, M., Mathur, P. & Guo, S. Corticotropin-releasing factor critical for zebrafish
4086 camouflage behavior is regulated by light and sensitive to ethanol. *The Journal of*
4087 *Neuroscience* **31**, 214-224 (2011).
4088 252 Nasiadka, A. & Clark, M. D. Zebrafish breeding in the laboratory environment. *ILAR*
4089 *Journal* **53**, 161-168 (2012).
4090 253 Marsh, D. J. *et al.* Response of melanocortin-4 receptor-deficient mice to anorectic
4091 and orexigenic peptides. *Nature Genetics* **21**, 119 (1999).
4092 254 Smith, C. J. & Haley, S. R. Steroid profiles of the female tilapia, *Oreochromis*
4093 *mossambicus*, and correlation with oocyte growth and mouthbrooding behavior.
4094 *General and Comparative Endocrinology* **69**, 88-98 (1988).
4095 255 Porter, D. T., Roberts, D. A. & Maruska, K. P. Distribution and female reproductive state
4096 differences in orexigenic and anorexigenic neurons in the brain of the mouth brooding
4097 African cichlid fish, *Astatotilapia burtoni*. *The Journal of Comparative Neurobiology* **525**,
4098 3126-3157 (2017).
4099 256 Field, K. E., McVicker, C. T. & Maruska, K. P. Sexually-Relevant Visual and Chemosensory
4100 Signals Induce Distinct Behaviors and Neural Activation Patterns in the Social African
4101 Cichlid, *Astatotilapia burtoni*. *Front Behav Neurosci* **12**, 267-267 (2018).
4102 257 Wootton, R. J. & Smith, C. *Reproductive biology of teleost fishes*. (John Wiley & Sons,
4103 2014).
4104 258 D'cotta, H., Fostier, A., Guiguen, Y., Govoroun, M. & Baroiller, J. Aromatase plays a key
4105 role during normal and temperature-induced sex differentiation of tilapia *Oreochromis*
4106 *niloticus*. *Molecular Reproduction and Development* **59**, 265-276 (2001).
4107 259 Shiau, A. K. *et al.* The structural basis of estrogen receptor/coactivator recognition and
4108 the antagonism of this interaction by tamoxifen. *Cell* **95**, 927-937 (1998).
4109 260 Vladusic, E. A., Hornby, A. E., Guerra-Vladusic, F. K., Lakins, J. & Lupu, R. Expression and
4110 regulation of estrogen receptor beta in human breast tumors and cell lines. *Oncology*
4111 *reports* **7**, 157-224 (2000).
4112 261 Singh, R., Singh, A. K. & Tripathi, M. Effect of a non steroidal tamoxifen on the gonad
4113 and sex differentiation in Nile tilapia, *Oreochromis niloticus*. *Journal of Environmental*
4114 *Biology* **33**, 799 (2012).
4115 262 El-Brolosy, M. A. *et al.* Genetic compensation triggered by mutant mRNA degradation.
4116 *Nature* **568**, 193-197 (2019).
4117
4118

4119 **Supplementary Data**

4120

Supplementary Table 1. Bodyweight to SNPs association summary in hybrid tilapia.

Genomic Position	Features	Allele	Individuals	Allele Frequency	Mean Bodyweight (220dph)(g)	220 dph p-val
chrLG1:28146922	<i>RASGRF1</i> Promoter	CT	122	0.772	350.9	0.027
		TT	36	0.228	315.7	
chrLG1:28146925	<i>RASGRF1</i> Promoter	CT	122	0.772	350.9	0.027
		CC	36	0.228	315.7	
chrLG1:28146966	<i>RASGRF1</i> Promoter	AG	122	0.772	350.9	0.027
		GG	36	0.228	315.7	

4121

4122

Supplementary Table 2. Microsatellite genotype of sex-markers linked to sex locus.

Individual	S1542		S1543		S001		S002		Sex
Father	356	382	378	407	337	339	383	383	Male
Mother	356	392	382	407	317	333	383	395	Female
1	356	356	407	407	333	339	383	383	Male
2	382	392	382	407	317	337	387	395	Female
3	382	392	382	407	317	337	387	395	Female
4	356	392	382	407	317	339	383	395	Male
5	356	356	407	407	333	339	383	383	Male
6	382	392	382	407	317	337	387	395	Female
7	356	392	382	407	317	339	383	395	Male
8	356	382	407	407	333	337	383	387	Female
9	356	382	407	407	333	337	383	387	Female
10	356	356	407	407	333	339	383	383	Male
11	382	392	382	407	317	337	387	395	Female
12	382	392	382	407	317	337	387	395	Female
13	356	356	407	407	333	339	383	383	Male
14	382	392	382	407	317	337	387	395	Female
15	382	392	382	407	317	337	387	395	Female
16	382	392	382	407	317	337	387	395	Female
17	356	382	407	407	333	337	383	387	Female
18	382	392	382	407	317	337	387	395	Female
19	356	382	407	407	333	337	383	387	Female
20	356	382	407	407	333	337	383	387	Female
21	356	392	382	407	317	339	383	395	Male
22	382	392	382	407	317	337	387	395	Female
23	382	392	382	407	317	337	387	395	Female
24	356	382	407	407	333	337	387	395	Female
25	356	392	382	407	333	339	383	383	Male
26	356	392	382	407	317	339	383	395	Male
27	356	356	378	407	333	339	383	383	Male
28	356	356	407	407	333	339	383	383	Male
29	356	392	382	407	317	339	383	395	Male
30	356	356	378	407	333	339	383	395	Male
31	382	392	382	407	317	337	387	395	Female
32	382	392	382	407	317	337	387	395	Female
33	382	392	382	407	317	337	387	395	Female
34	356	382	378	407	333	337	383	383	Female
35	356	382	378	407	333	337	383	383	Female
36	356	382	378	407	333	337	387	395	Female
37	356	392	378	382	317	339	383	395	Male
38	356	392	382	407	317	339	383	395	Male
39	382	392	382	407	317	337	387	395	Female
40	356	382	407	407	333	337	383	387	Female
41	356	382	407	407	333	337	383	387	Female

42	356	356	407	407	333	339	383	383	Male
43	356	356	407	407	333	339	383	383	Male
44	356	382	407	407	333	337	383	387	Female
45	382	392	407	407	317	337	387	395	Female
46	356	356	378	407	333	339	383	383	Male
47	356	382	407	407	333	337	383	387	Female
48	382	392	407	407	317	337	387	395	Female
49	382	392	407	407	317	337	387	395	Female
50	356	356	378	407	333	339	383	383	Male
51	356	356	378	407	333	339	383	383	Male
52	356	356	378	407	333	339	383	383	Male
53	356	356	378	407	333	339	383	383	Male

4123

Correlation between RNA-seq and RT-qPCR from Muscle Transcriptome

Supplementary Figure 1. Scatter plot between log₂ fold-change in muscle gene expression from qPCR verfications and RNA-seq analysis. Linear solid line showed best linear fit between qPCR log₂ Fold-change and RNA-seq log₂ fold-change. Pairs of dotted line showed 95% confidence interval for correlation between qPCR log₂ Fold-change and RNA-seq log₂ fold-change. The Pearson R is 0.9117 with a non-zero correlation *p*-val of 0.01.

Correlation between RNA-seq and RT-qPCR from Brain Transcriptome

4132

4133 Supplementary Figure 2. Scatter plot between log₂ fold-change in brain gene expression from qPCR verifications
4134 and RNA-seq analysis. Linear solid line shows best linear fit between qPCR log₂ Fold-change and RNA-seq log₂
4135 fold-change. Pairs of dotted line shows 95% confidence interval for correlation between qPCR log₂ Fold-change
4136 and RNA-seq log₂ fold-change. The Pearson R is 0.9304 with a non-zero correlation *p*-val < 0.001.

4137