Genome-wide search for quantitative trait loci (QTL) and genes for increased growth and improved fatty acid content in salt tolerant tilapia (oreochromis spp.)

Lin, Grace Xiulin

2018

Lin, G. X. (2018). Genome-wide search for quantitative trait loci (QTL) and genes for increased growth and improved fatty acid content in salt tolerant tilapia (oreochromis spp.). Doctoral thesis, Nanyang Technological University, Singapore.

<http://hdl.handle.net/10356/74843>

<https://doi.org/10.32657/10356/74843>

**GENOME-WIDE SEARCH FOR QUANTITATIVE
TRAIT LOCI (QTL) AND GENES FOR
INCREASED GROWTH AND IMPROVED FATTY
ACID CONTENT IN SALT TOLERANT TILAPIA
(*OREOCHROMIS SPP.*)**

**LIN XIULIN, GRACE
SCHOOL OF BIOLOGICAL SCIENCES
2017**

GENOME-WIDE SEARCH FOR QUANTITATIVE
TRAIT LOCI (QTL) AND GENES FOR INCREASED
GROWTH AND IMPROVED FATTY ACID CONTENT
IN SALT TOLERANT TILAPIA (*OREOCHROMIS SPP.*)

Lin Xiulin, Grace

School of Biological Sciences

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

2017

Acknowledgements

I would like to thank my supervisors, Dr. Gen Hua Yue, Professor Chun Ling, Valerie Lin and Professor László Orbán for their guidance, constructive criticisms, and kind support throughout my Ph.D. candidature. With their insightful advice, I was able to gain more knowledge on molecular and population genetics and become a better researcher. I would also like to thank Dr. Mark Featherstone and my other advisory committee members, Drs Vladimir Korzh and Li Feng Zhang for their kind support and help during my PhD journey.

I am thankful to my colleagues: Drs Feng Liu, Le Wang, Keyi Ma for their technical expertise and insightful exchanges. Thanks to Ms May Lee, Ms Minghui Liao and Ms Elaine Chua for their help with some of my experiments. Special thanks to Ms Natascha May Thevasagayam and Mr Prakki Sai Rama Sridatta for their bioinformatics support. Their friendship and help made my PhD journey much more rewarding.

I am also grateful for the generous funding from Temasek Life Sciences Laboratory, Agri-Food & Veterinary Authority of Singapore and National Research Foundation.

Most importantly, I would like to thank my family, especially my husband and my parents for their understanding, encouragement and loving support throughout all these years.

Table of Contents

Acknowledgements -----	i
List of figures -----	vii
List of tables -----	ix
List of appendices -----	x
List of publications and presentations -----	xi
Abbreviations -----	xii
Abstract-----	xv
Chapter 1. Introduction-----	1 -
1.1 Current status of conventional breeding of aquaculture-----	1 -
1.2 Molecular breeding concept, advantages, requirements, and status -----	3 -
1.2.1 Establishing mapping populations and phenotyping -----	4 -
1.2.2 Genetic markers -----	5 -
1.2.2.1 Traditional genetic markers and genotyping -----	5 -
1.2.2.2 Genotype by sequencing (GBS) -----	10 -
1.2.3 Construction of linkage maps-----	13 -
1.2.4 Quantitative trait loci (QTL)-----	16 -
1.2.4.1 QTL mapping-----	17 -
1.2.4.2 Gene-based/Candidate gene approach in mapping traits-----	22 -
1.3 Growth of fishes and the candidate genes-----	27 -
1.3.1 Factors involved in growth regulation of fishes-----	27 -
1.3.2 Myotomal growth in fishes-----	29 -
1.3.3 Sexual dimorphism and growth -----	30 -
1.3.4 Status of mapping QTL and transcriptome analysis of growth in aquaculture species -----	31 -
1.4 Meat quality: Lipid content of fishes -----	33 -
1.4.1 Total Lipid content in fishes -----	33 -
1.4.2 Biosynthesis and metabolism of fatty acids-----	35 -

1.4.3 Mapping QTL and transcriptome analysis for lipid composition in aquaculture species -----	38 -
1.5 Tilapia biology and genetics-----	40 -
1.5.1 Biology -----	40 -
1.5.2 Genetic improvements via selective breeding -----	41 -
1.5.3 Genomic tools in tilapia-----	43 -
1.5.3.1 Molecular markers and linkage mapping -----	43 -
1.5.3.2 Genome sequencing and transcriptomics -----	43 -
1.5.3.3 QTL mapping for economical traits-----	45 -
1.6 Aim and outline of this thesis -----	48 -
1.7 Project roadmap-----	51 -
Chapter 2. Materials and Methods -----	52 -
2.1 Ethics statement-----	52 -
2.2 Mapping population, salinity acclimatization, and fish husbandry -----	52 -
2.3 Measuring growth traits and collection of fin clips and dorsal muscle tissue---	53 -
2.4 Visual sexing of tilapia -----	54 -
2.5 Determining total neutral lipid content and profiling fatty acids -----	55 -
2.5.1 Extraction and quantification of total neutral lipid -----	55 -
2.5.2 FAME extraction and quantification and qualification of fatty acids-----	56 -
2.6 Preparation of genomic DNA -----	57 -
2.6.1. Genomic DNA isolation for QTL scan for growth and sex -----	57 -
2.6.2. Genomic DNA isolation for QTL mapping for fatty acid contents-----	58 -
2.7 Preparation of RNA -----	60 -
2.7.1 RNA isolation and assessment of quantity and quality -----	60 -
2.8 QTL mapping - Genotyping -----	62 -
2.8.1 Genotyping of microsatellites for QTL for growth and sex-----	62 -
2.8.2 Genotype by sequencing (GBS) library preparation -----	62 -
2.8.3 Processing of NGS reads, identifying and genotyping SNPs for scan of QTL for growth and fatty acids-----	63 -
2.9 Construction of linkage maps -----	64 -

2.9.1 Linkage map construction using microsatellites -----	64 -
2.9.2 Linkage map construction using microsatellites and SNP -----	65 -
2.10 QTL mapping -----	65 -
2.10.1 QTL mapping for growth traits and sex -----	65 -
2.10.1.1 Fine mapping of significant QTL for growth in male offspring -----	66 -
2.10.2 QTL mapping for total lipid and fatty acid content (Omega 3 and 6)-----	67 -
2.11 Transcriptomic analysis of fast and slow growing fish -----	67 -
2.11.1 RNA-Sequencing library preparation-----	67 -
2.11.2 Processing of sequencing reads and <i>de novo</i> assembly of reference transcriptome-----	67 -
2.11.3 Identification of differentially expressed transcripts in the muscle and liver --	69 -
2.11.4 Functional classification and pathway analysis of differentially expressed transcripts (DETs)-----	69 -
2.11.5 Reverse transcription to convert RNA to single-stranded cDNA (sscDNA)---	70 -
2.11.6 Real-time quantitative reverse transcriptase-PCR (qRT-PCR) for validation of expression of differentially expressed transcripts (DETs) -----	71 -
2.12. Differentially expressed genes with sequence variation that are located within or with close proximity of the QTL regions for growth and fatty acid content-----	72 -
2.12.1 Genomic location of DEGs with identified QTL regions for growth and fatty acid content through linkage and physical mapping-----	72 -
2.12.2 Tissue expression of DEGs -----	73 -
2.13 Statistical analyses -----	73 -
Chapter 3. Results -----	74 -
3.1 Trait values and their correlations -----	74 -
3.1.1 Growth traits -----	74 -
3.1.2 Total neutral lipid and fatty acid content -----	76 -
3.1.2.1 Correlations between fatty acids-----	76 -
3.1.2.2. Correlation between fatty acids and growth traits -----	77 -
3.2 Linkage maps -----	80 -

3.2.1 Construction of linkage maps for all offspring, male offspring, female offspring as well as male and female offspring -----	80 -
3.2.2 Comparison between male and female linkage maps -----	81 -
3.2.3 Construction of a denser linkage map using GBS technology -----	83 -
3.2.3.1 Identification and genotyping of genome-wide SNPs -----	83 -
3.2.3.2. Construction of denser linkage map using microsatellite and SNP markers -----	84 -
3.3 Mapping QTL for economical traits -----	92 -
3.3.1 Mapping QTL using the microsatellite-based linkage map -----	92 -
3.3.1.1 QTL for sex -----	92 -
3.3.1.2 Growth in all offspring, only male, and only female offspring -----	93 -
3.3.2 Improved QTL mapping using the microsatellite and SNP based linkage map -	99 -
3.3.2.1 QTL for growth traits -----	99 -
3.3.2.2 QTL for total lipid and fatty acid content -----	102 -
3.4 Transcriptome analysis of fast and slow growing fish -----	107 -
3.4.1 Raw sequencing reads and processing -----	107 -
3.4.2 <i>De novo</i> assembly of reference transcriptome -----	107 -
3.4.3 Differential expression of transcripts between the muscle and liver of fast and slow-growing fish -----	109 -
3.4.4 Functional classification of DETs -----	111 -
3.4.5 Validation of differentially expressed genes (DEG) using quantitative PCR (qPCR) transcripts -----	113 -
3.4.6 Selected KEGG pathways associated with muscle development and lipid biosynthesis and the DEGs in muscle and liver -----	116 -
3.5 Differentially expressed genes (DEG) with sequence variation that are located inside or within close proximity of the QTL -----	116 -
3.5.1 Genomic locations of DEG -----	116 -
3.5.2 Tissue expression of DEGs in muscle and liver -----	119 -
Chapter 4. Discussion -----	120 -

4.1 Construction of first densest linkage map of saline tilapia using both microsatellites and SNP markers -----	120 -
4.2 Improved growth performance of saline tilapia-----	125 -
4.3 Increased neutral lipid content and Omega 3 fatty acid of saline tilapia-----	126 -
4.4 Increased growth traits in saline tilapia associated with multiple genomic regions - -	130 -
4.5 Improved omega 3 fatty acid content in saline tilapia is associated with four genomic regions -----	135 -
4.6 Increased growth and improved lipid contents are associated with Hippo pathway, decrease in apoptosis and lipogenesis in fast-growing fishes-----	138 -
4.7 Summary and future directions -----	145 -
References -----	148 -

List of figures

Figure 1.1. Global aquaculture production of farmed fish.	1 -
Figure 1.2. Principle of genotype by sequencing (GBS)	12 -
Figure 1.3. Recombination events between homologous chromosomes during meiosis.....	14 -
Figure 1.4. The principle of QTL mapping.	18 -
Figure 1.5. Three of the widely used QTL mapping analysis.....	21 -
Figure 1.6. A simplified overview of RNA-seq and data analysis.	26 -
Figure 1.7. Most common types of storage and membrane lipids.	34 -
Figure 1.8. Biosynthetic pathway of very long chained polyunsaturated fatty acids (VLC-PUFA).- 36	-
Figure 1.9. Flowchart of experimental design for the discovery of QTL, candidate genes and genetic markers for growth and fatty acid content in salt tolerant tilapia.	51 -
Figure 2.1. Morphometric measurements obtained from F ₂ mapping population.....	54 -
Figure 2.2. Genital papilla of tilapia.	55 -
Figure 2.3. Preparation of samples for lipid and fatty acid extraction through freeze-drying and grinding of the dorsal muscle of tilapia.	56 -
Figure 2.4. Flowchart of transcriptome assembly and annotation of salt tolerant tilapia.....	68 -
Figure 2.5. Frequency distribution of reads per exon kilobase per million mapped reads (RPKM) values of differentially expressed transcripts in dorsal muscle and liver of salt tolerant tilapia. ..	70 -
Figure 3.1. Frequency distribution of three growth traits in salt tolerant tilapia.	75 -
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring (continued). - 86	-
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring (continued)- 87 -	-
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring (continued)- 88 -	-
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring (continued)- 89 -	-
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring (continued)- 90 -	-
Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F ₂ offspring.....	91 -
Figure 3.3. QTL for sex detected on LG 1 and LG 22 in salt-tolerant tilapia.	92 -
Figure 3.4. QTL for growth traits detected on LG 2 in male offspring of salt-tolerant tilapia.	96 -
Figure 3.5. QTL for growth traits detected on LG 18 in male offspring of salt-tolerant tilapia....	97 -
Figure 3.6. QTL for growth traits detected on LG 8 in female offspring of salt-tolerant tilapia...-	98 -
Figure 3.7. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 18 in salt tolerant tilapia.....	103 -
Figure 3.8. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 20 in salt tolerant tilapia.....	104 -
Figure 3.9. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 11 in salt tolerant tilapia.....	105 -
Figure 3.10. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 19 in salt tolerant tilapia.....	106 -
Figure 3.11. Length distribution of contigs from the <i>de novo</i> assembled reference transcriptome of salt tolerant tilapia.....	109 -
Figure 3.12. Differentially expressed transcripts (DET) in the muscle and liver of salt tolerant tilapia at 110 dph when comparing between fast to slow-growing fishes.	110 -

Figure 3.13. Top 10 functional classification of differentially expressed transcripts in the muscle (Top) and liver (Bottom).	 - 112 -
Figure 3.14. Top 20 KEGG pathways of differentially expressed transcripts in the muscle and liver of salt tolerant tilapia.	 - 113 -

List of tables

Table 1. Summary of DNA markers, their characteristics, potential applications and limitations .-	9 -
Table 3.1. Averages growth traits for all, male and female offspring recorded in 180 dph (days post-hatch) in an F2 full-sib mapping family of salt-tolerant tilapia	75 -
Table 3.2. Correlations between recorded growth traits of salt tolerant tilapia at 180 dph.....	75 -
Table 3.3. Phenotypic values of total neutral lipid (mg/g), fatty acid content (mg/g wet weight), fat and moisture content in the flesh of 246 salt tolerant tilapia individuals.....	78 -
Table 3.4. Correlations between fatty acid compositions of salt tolerant tilapia (356 individuals) at 220 dph	79 -
Table 3.5. Correlations between fatty acid and growth traits of salt tolerant tilapia	80 -
Table 3.6. Summary of genetic linkage map of all offspring, male offspring, female offspring and male and female offspring in salt tolerant tilapia	82 -
Table 3.7. Recombination ratios between male and female genetic maps in salt tolerant tilapia ..	83 -
Table 3.8. Summary of statistics for the linkage map constructed using microsatellite and SNP markers for QTL mapping in salt tolerant tilapia; total length, number of markers and average marker distance	85 -
Table 3.9. Significant and suggestive quantitative trait loci (QTL) for growth traits and sex in all, male and female offspring of salt-tolerant tilapia	94 -
Table 3.10. Significant and suggestive quantitative trait loci for growth traits of salt tolerant tilapia -	100 -
Table 3.11. Significant and suggestive quantitative trait loci for total lipids, total fatty acids and fatty acid content in salt tolerant tilapia.....	101 -
Table 3.12. Statistics from sequencing reads of ten libraries generated for saline tilapia using Illumina HiSeq2000 sequencing	107 -
Table 3.13. Overview of reference transcriptome assembled using different software for RNA-seq sequences generated from ten libraries	108 -
Table 3.14. Mapping efficiency and coverage of pair-end libraries of the muscle and liver of the fast and slow growing fishes	110 -
Table 3.15. Fold changes of selected genes determined by qPCR for validation of RNA-seq	115 -
Table 3.16. Differentially expressed genes that lie within/close to the QTL region of associated traits of F2 hybrid saline tilapia	118 -
Table 3.17. Average gene expression relative to actin genes analyzed by qPCR, standard deviations and significance of DEGs that lie within the QTL regions in dorsal muscle and liver of fast and slow growing F2 hybrid salt tolerant tilapia.....	119 -

List of appendices

Appendix 1. Selected microsatellite markers for reconstruction of a linkage map in salt-tolerant tilapia	171
Appendix 2. Fine-mapping microsatellite marker pairs for male linkage group	- 179 -
Appendix 3. List of primers used for real-time qPCR analyses of gene transcript analysis for the verification of differentially expressed genes in the muscle and liver of salt tolerant tilapia.	- 180 -
Appendix 4. A genetic linkage map of salt-tolerant tilapia constructed using all offspring.	- 181 -
Appendix 5. A genetic linkage map of salt-tolerant tilapia constructed using only male offspring.	- 182 -
Appendix 6. A genetic linkage map of salt-tolerant tilapia constructed using only female offspring...	- 183 -
Appendix 7. A genetic linkage map of salt-tolerant tilapia constructed using male and female offspring.	- 184 -
Appendix 8. QTL profile for growth traits and sex determination in all offspring.	- 185 -
Appendix 9. QTL profile for growth traits in male offspring.	- 186 -
Appendix 10. QTL profile for growth traits in female offspring.	- 187 -
Appendix 11. Differentially expressed genes in the muscle in selected KEGG pathways.	- 188 -
Appendix 12. Differentially expressed genes in the liver of selected KEGG pathways.	- 193 -
Appendix 13. List of primers used for gene expression studies of candidate genes in muscle and liver of fast and slow- growing fishes.	- 205 -

List of publications and presentations

Journal publications

Xia JH, **Lin G**, He X, Liu P, Sun F, Tu R & Yue GH. (2013) Whole genome scanning and association mapping identified a significant association between growth and an SNP in the IFABP-a gene of the Asian seabass. BMC Genomics 14:295

Xia JH, **Lin G**, Fu GH, Wan ZY, Lee M, Wang L, Liu, XJ, Yue GH. (2014) The intestinal microbiome of fish under starvation. BMC Genomics 15: 266

Xia JH, **Lin G**, He XP, Bu YP, Liu P, Liu P, Sun F, Tu RJ and Yue GH. (2014) Mapping Quantitative Trait Loci for Omega-3 Fatty Acids in Asian Seabass. Marine Biotechnology 16: 1-9.

Wan ZY, Xia JH, **Lin G**, Wang L, Lin VCL & Yue GH. (2016) Genome-wide methylation analysis identified sexually dimorphic methylated regions in hybrid tilapia. Scientific Reports 6: 35903

Lin G, Chua E, Orbán L and Yue GH. (2016) Mapping QTL for sex and growth traits in salt-tolerant tilapia (*Oreochromis spp.* X *O. Mossambicus*). PLOS One 11: e0166723.

Lin G, Wang L, Ngho ST, VCL Lin, Orbán L and Yue GH. Mapping QTL for Omega-3 Content in Hybrid Saline Tilapia. Marine Biotechnology 20 (1): 10-19

Lin G, Thevasagayam N.M, VCL Lin, Orbán L and Yue GH. Transcriptome survey of gene expression profiles underlying growth superiority and lipid biosynthesis of adult hybrid salt tolerant tilapia (*Marine Biotechnology under review*).

Conference presentations (presenter underlined)

Lin G., Yue G.H (2015). Genomics: Marker-assisted selection. Genome-wide search for QTL for growth traits of saline tilapia (*Oreochromis spp.*) International Symposium on Genetics in Aquaculture XII (ISGA) 21st -27th June. Santiago de Compostela, Spain (Poster).

Abbreviations

AFLP	Amplified Fragment Length Polymorphism
ALA	Alpha Linolenic Acid
bp	Basepairs
BP	Biological Processes
BT	Body Thickness
BW	Body Weight
C**	Carbon_numericals
CC	Cellular Component
cDNA	Complementary DNA
CDS	Coding DNA Sequences
CIM	Composite interval mapping
cM	Centimorgan
CRoPs	Complexity Reduction of Polymorphic Sequences
CS	Suggestive at Chromosome-wide
DEG	Differentially Expressed Gene (s)
DET	Differentially Expressed Transcript (s)
DH	Double Haploids
DHA	Docosahexaenoic Acid
DNA	Deoxyribonucleic Acid
DPH	Days Post Hatch
dsDNA	double-stranded DNA
EPA	Eicosapentaenoic Acid
EST	Expressed Sequence Tag
FA	Fatty Acids
FAME	Fatty Methyl Esther
GBS	Genotype By Sequencing
GC-MS	Gas Chromatograph Mass Spectrophotometry
gDNA	Genomic DNA
GO	Gene Ontology
GS	Genomic Selection
GWAS	Genome-wide Associated Selection
HUFA	Highly Unsaturated Fatty Acids
Indel	Insertions or deletions
KAAS	KEGG Automatic Annotation Server

Kbp	Kilo base pairs
KEGG	Kyoto Encyclopaedia of Genes and Genomes
LD	Linkage Disequilibrium
LG	Linkage Group
LOD	Logarithm of Odds
MAS	Marker Assisted Breeding
Mbp	Million basepairs
MF	Molecular Function
miRNA	MicroRNA
ML	Maximum Likelihood
MQM	Multiple QTL mapping
mRNA	Messenger RNA
MS	Microsatellites
MSG	Multiplexed Shotgun Genotyping
MUFA	Monounsaturated Fatty Acids
n3	Omega 3 fatty acids
n6	Omega 6 fatty acids
ncRNA	Non-coding RNA
NGS	Next Generation Sequencing
NILs	Near Isogenic Lines
PCR	Polymerase Chain Reaction
PE	Paired End
PPP	Pentose Phosphate Pathway
PPT	Parts Per Thousand
PUFA	Polyunsaturated Fatty Acids
qPCR	Quantitative Polymerase Chain Reaction
QTL	Quantitative Trait Locus/loci
QTL	Quantitative trait
RAD-seq	Restriction site Associated DNA sequencing
RAPD	Random Amplification of Polymorphic DNA
RFLP	Restriction Fragment Length Polymorphism
RIN	RNA Integrity Number
RNA	Ribonucleic Acid
RNA-seq	RNA sequencing
RPKM	Reads Per exon Kilobase per Million mapped reads

RRL	Reduced Representation Library
rRNA	Ribosomal RNA
SC	Significant at Chromosome-wide
SE	Single End
SFA	Saturated Fatty Acids
SG	Genome-wide Significance
SIM	Simple interval mapping
SNP	Single Nucleotide Polymorphism
sscDNA	Single stranded cDNA
TCA	Tricarboxylic Acid cycle
TDW	Total lipid (Dry Weight)
TFA	Total Fatty Acids
TL	Total Length
tRNA	Transfer RNA
TWW	Total lipid (Wet Weight)
UTR	Untranslated Regions
WGD	Whole genome duplication

Abstract

Tilapias are the second most important group of farmed fish in the world. Increasing demands for fish as a protein source as well for heart health and decreasing land space due to the growing population has highlighted the importance of breeding fast growing salt tolerant tilapia strains/hybrids with high omega 3 fatty acids for aquaculture. Prior to this project, there was little information about growth and omega 3 fatty acid content in saline tilapia as previous studies were mainly based on freshwater tilapia. This Ph.D. project aims to conduct quantitative trait loci (QTL) mapping for growth traits and fatty acid profiles and to identify for candidate genes located in significant QTL using RNA sequencing in salt tolerant tilapia to accelerate the genetic improvement of these two important traits. It also seeks to improve the understanding of the mechanisms underlying phenotypic variation of growth traits and fatty acid profiles.

In the first part of the thesis, QTL mapping for growth, sex and fatty acid content have been conducted in an F₂ family of salt tolerant tilapia with 522 individuals. To conduct QTL mapping, linkage maps were constructed. Genotype by sequencing (GBS) technology was utilised to discover and genotype Single Nucleotide Polymorphism (SNP) markers, and together with microsatellites, the first densest linkage map with 784 markers, consisting of 22 linkage groups for saline tilapia was constructed. Phenotyping of growth and fatty acid traits were also carried out. The growth of the salt-tolerant tilapia was comparable to Nile tilapia that were selected for increased growth. The omega 3 fatty acid content of salt tolerant tilapia was much higher than that in freshwater tilapia and was comparable to that of fatty fishes such as mackerel and chinook salmon. QTL mapping was first conducted using the microsatellite-based linkage map and showed that salt tolerant tilapia had two sex associated genomic regions on LG 1 and LG 22, explaining 19.8% and 10.8% of the phenotypic variance respectively. Additionally, growth-associated regions in male and female offspring were generally different; for male offspring, QTL were located on LG 2 and 18, explaining 5.1% to 10.1% and 8.5% to 10.2% of phenotypic variance, respectively and for female offspring, QTL was located

on LG 8 and explained 10.4% to 19.9% of phenotypic variance in the population. To improve resolution on the QTL for growth and omega 3 fatty acids, the microsatellite and SNP marker based linkage map was utilized for mapping. This QTL mapping detected two significant QTL for growth on LG 2 and 6, explaining 5.1% and 3.6% to 5.2% of phenotypic variance, respectively. For total lipid and fatty acid content, one genome-wide QTL detected on LG 8, explaining 5.6 to 8.3% phenotypic variance was associated with four traits and 14 significant QTL detected on six LG, namely LG 4, 13, 15, 16, 18 and 20, explaining 6.5% to 17.0% of phenotypic variance were associated with nine traits in the population. The markers flanking the QTL for growth and omega 3 fatty acids can be used in MAS to accelerate the genetic improvements of salt tolerant tilapia for these traits.

For the second part of the thesis, a reference transcriptome was generated and differentially expressed genes in the dorsal muscle and liver between fast and slow growing tilapia were identified using RNA sequencing. Pathway analysis using Kyoto Encyclopedia of Genes and Genomes (KEGG) was conducted to classify these genes into different pathways. From the differentially expressed genes identified, ten genes are found to be located within or at close proximity to QTL detected for growth and omega 3 fatty acid content in salt tolerant tilapia. These ten genes could be candidate genes for growth and omega 3 fatty acid content. In addition, qRT-PCR was conducted and verified the differential expression of these ten genes in the dorsal muscle and liver between fast and slow growing fishes.

To summarize, the data obtained through QTL mapping and RNA sequencing clearly suggest the multitude of genomic regions and/or genes that may underlie the phenotypic variations in growth and omega 3 fatty acids in salt tolerant tilapia. Nonetheless, this study lays the groundwork for further investigation of the molecular mechanisms and underlying cause of the phenotypic variation of these two traits. Markers tightly linked to the QTL for growth and omega-3 content could be used in

marker-assisted selection (MAS) programmes to accelerate the improvement of growth performance and meat quality of salt tolerant tilapia.

Chapter 1. Introduction

1.1 Current status of conventional breeding of aquaculture

On earth, there are more than 30,000 teleost fish and hundreds of thousands of shellfish species (Nelson *et al.* 2006) however, only approximately 600 aquatic species are being farmed globally and they include finfish, crustaceans, molluscs, amphibians and other aquatic reptiles and animals such as crocodiles, sea squirts and sea cucumbers. Of which carps, tilapia and oysters have the largest worldwide production (FAO 2016). Aquaculture or fish farming is one of the fastest growing primary industries in the world with global production of more than 70 million tonnes in 2014 and this provides more than half the supplies destined for human consumption (Figure 1.1). So far in aquaculture, only 10% of the world's production utilized genetically improved stocks (Sonesson & Meuwissen 2009) and this meant that a huge proportion of the remaining aquaculture productions are still utilizing traditional domestication and/or are at an early stage of developing improvements.

Figure 1.1. Global aquaculture production of farmed fish.

Adapted from (FAOSTAT 2017)

Conventional breeding for genetic improvement is based on phenotypic measurements and the estimation of genetic parameters with quantitative genetic

approaches. One of the first documented selection in fishes was carried out in brook trout (*Salvelinus fontinalis*) for disease resistance against furunculosis (Embry & Hayford 1925), increasing survival rate to 69% merely over three generations (Gjedrem 2012). One of the pioneering selection programmes conducted in Atlantic salmon (*Salmo salar*) has reported an average of 14% genetic gain per generation (Gjedrem 2010) from 1975 until 2010. Since then, more than forty selective breeding programmes in aquatic species have been initiated in Asia, Africa and Latin America alone, mainly focused on growth-related traits (Neira 2010).

Conventional breeding are mainly based on traits that can be observed physically and/or genetically linked such as growth rate, gender, age at maturation, feed conversion efficiency, nutritional value and cold, salinity and disease resistance, which can be affected by non genetic factors such diet, seasons, post harvesting and post slaughtering procedures to genetically linked traits (Beaumont *et al.* 2010). However, the improvement of these traits through conventional breeding means may prove to be a laborious and of low efficiency (Gjedrem & Baranski 2009; Nguyen *et al.* 2010a). This is especially the case with traits that are difficult to measure or observe physically but have major effects of the profitability and productivity of the species such as disease resistance, feed conversion rate (FCR), fatty acid profiling and flesh quality. Additionally, the selection of these traits of interest is affected by environmental factors (e.g. feed, water quality, culture density). Even though conventional breeding schemes have been successful in domesticated animals and even in plants (Gjedrem 2010), the genetic improvement of aquatic species is still at the developmental stage. Molecular breeding or marker-assisted selection (MAS) based on DNA markers, which are tightly linked to traits and can precisely predict traits, can solve some problems in conventional breeding and thus, accelerate genetic improvement.

1.2 Molecular breeding concept, advantages, requirements, and status

Molecular breeding is a generic term used to describe the application of molecular tools for utility in breeding programmes. With the advent of molecular techniques such as next-generation sequencing (NGS), utilising molecular tools may shorten the time and increase the productivity of fish breeding through sieving out of quality phenotypes in an effective and efficient manner. The concept of MAS for aquatic species began in the late 1900s (Gjedrem & Baranski 2009), utilizing variation in selected DNA sequences as markers and associating phenotypic data to these markers in an attempt to improve selection response (Meuwissen & Goddard 1996). The main purpose of MAS is to combine genomic information at the DNA markers with the trait and conduct selection based on markers linked to the trait instead of the trait itself (Ribaut & Ragot 2007). This meant that the effect of the genes/markers is measured directly through the genetic makeup of the animal, regardless of the environmental conditions, rather than the estimation through phenotype, thus allowing for faster selection processes. Therefore, using MAS to improve selection response is more advantageous than conventional breeding since it can be used on traits that are difficult to measure, expressed late in development and/or exhibit low heritabilities (Lande & Thompson 1990; Meuwissen & Goddard 1996). Since MAS uses the absence/presence of a marker to assist in phenotypic selection, identifying the specific polymorphisms responsible for the phenotype is required (Collard *et al.* 2005). One of the major limitations in MAS is the cost of sampling and genotyping these samples. However, with advancing genomic and genotyping technologies, the use of MAS in breeding programmes is increasing. In MAS, the identification of quantitative trait loci (QTL) is one of the essential steps. There are three essential requirements in the mapping of QTL. Firstly, the mapping population, where both DNA markers and phenotypes segregate, must be established and the phenotyping of this mapping population must be carried out. Second, the development

of genetic markers and genotyping them as they will be utilised for the construction of a genetic/linkage map. Lastly, the analysis of the locations and effects of the QTL using statistical software to associate the genotype with the phenotype at specific locations across the genome.

1.2.1 Establishing mapping populations and phenotyping

Mapping populations are crucial components of QTL mapping as they play a major role in the construction of a genetic/linkage map. Aquatic species have high fecundity and commonly outbreed, avoiding inbreeding due to negative effects (Kincaid 1983; Frommen & Bakker 2006). Therefore, ideal mapping families such as near isogenic lines (NILs) and double haploids (DHs), like those in self-pollinating plants, are lacking due to the cost and time required, especially in species with long generation time. The high fecundity of fishes as previously mentioned hold a large resource, producing large family sizes. Furthermore, the outbreeding nature of fishes ensures heterozygosity in the population, which compensates for the loss of power of QTL mapping due to the lack of inbred lines (Paterson 1997). Therefore, the heterozygosity that comes parental lines will create segregations of genotypes and phenotypes in the mapping populations, generated either from F_1 , F_2 (cross of F_1 male and female) and backcross (cross of F_1 to one parent), are suitable for QTL mapping in aquaculture species (Paterson 1997; Yue 2014). The heterozygosity in the parental lines need not necessarily be phenotypic in nature, but genetically the alleles for the desired trait needs to be different so reassortment due to recombination will generate offspring populations that produce a range of phenotypic values (Miles & Wayne 2008). Once the type of mapping population for that particular trait and species is determined, the size of the mapping population is the next key consideration. Preliminary studies on QTL mapping requires a population size of 100-250 individuals (Mohan *et al.* 1997). However, larger populations will be required for higher resolution mapping and/or if the genome size and complexity of the species

increases. This is because a larger population size enables more recombinants to be generated and increases the statistical power, estimates of gene effects and confidence levels in the location of the QTL (Beavis 1998; Collard *et al.* 2005).

1.2.2 Genetic markers

Genetic markers are variable regions of DNA sequences that are used to track a particular location on a particular chromosome. They are heterozygous and facilitate the study of the inheritance of a gene or a trait hence, crucial for the construction of linkage maps. Thus, various technologies have been applied to aid the discovery and genotyping of these genetic markers especially for aquatic species (Liu & Cordes 2004).

1.2.2.1 Traditional genetic markers and genotyping

The advances in DNA marker technology made it possible to observe and exploit genetic variation in the genome. Different types of markers have different characteristics in terms of implementation and utility. Popular genetic markers commonly used for aquatic species include allozymes, mitochondrial DNA, restriction fragment length polymorphism (RFLP) (Botstein *et al.* 1980), random amplification of polymorphic DNA (RAPD) (Williams *et al.* 1990), amplified fragment length polymorphism (AFLP) (Vos *et al.* 1995), microsatellite (Weber & May 1989; Bruford & Wayne 1993), SNP (Vignal *et al.* 2002), expressed sequence tag (EST) markers (Schuler *et al.* 1996) and insertion or deletions (indels) (Table 1). Even though three of the marker systems, AFLP, microsatellite and SNP are able to produce a high number of polymorphisms, the most popular markers used for aquatic species are still microsatellites and SNPs. This is because they are able to be utilized for high throughput studies and produce a large number of polymorphisms with high reproducibility.

Microsatellites are tandemly arranged simple sequence repeats with a repeat number of 2-6 base pairs (bp). They are estimated to be present one in every 10 kilobase

(kb) in fish species and appear to be evenly distributed across all genomes (Liu & Cordes 2004). Their presence can be found in both regulatory and gene coding regions including promoters, enhancers, silencers, untranslated regions (UTRs), exons and introns. The function of the microsatellites can include the regulation of the expression of traits through its direct or indirect influence on gene transcription, splicing, transcript stability, degradation and transportation and translation efficiency (Yue 2014; Vieira *et al.* 2016). The major advantage of microsatellites is the multi-allelic nature, due to the variable number of repeats which, allows for the highest polymorphic information content to be available when compared to other genetic markers (Liu & Cordes 2004; Gjedrem & Baranski 2009). Nonetheless, the development of microsatellite markers requires a significant amount of resources in order to apply them for population studies including parentage assignment, kinship mapping, population structure, product traceability and genome mapping. This is because genomic segments containing microsatellites have to be enriched and their flanking regions must be sequenced for the design of primer pairs targeting the region. Additionally, the introduction of fluorescently labelled primers for automated scoring software like GeneMapper from ABI significantly increases the throughput and reproducibility of large-scale studies and cost per sample genotyped. An advantage with these predesigned primer pairs from microsatellites is that they can be cross-amplified to close species thereby reducing developmental costs. A few examples have been reported in aquatic species such as Asian green mussels (Lin *et al.* 2007), red coral grouper (Zhu *et al.* 2005) and hawksbill turtle (Lin *et al.* 2008). All in all, the highly polymorphic and co-dominant nature of microsatellites together with its high reproducibility and utility in large-scale studies have become one of the standards in genetic studies in a variety of fish species (Liu & Cordes 2004; Yue 2014). In spite of that, one of the major problems associated with the use of microsatellites is that there can be relatively high error rates since most genotyping and allele-calling procedures are semi-automated (Pompanon *et al.* 2005). The human-based error can be introduced due to the fluorescent labeling of the microsatellite

markers. These errors can be due to modification of PCR products and may influence the migration of the products (Hahn *et al.* 2001). Even though the errors can be circumvented, they are rarely conducted (Ginot *et al.* 1996) hence, inconsistencies amongst data are present due to the large variety of automatic sequencing machines, fluorescent dyes, and allele calling software used by different laboratories.

In the recent years, SNP markers have been utilized due to their low error rates and their simplicity in allele calling where the alleles are simply scored as present or absent (Vignal *et al.* 2002). SNPs are variations in a single nucleotide that occurs at a specific position in the genome (Sachidanandam *et al.* 2001) and are the most abundant type of DNA sequence polymorphism. In principle, any four possible nucleotides of A, T, G, and C can be present at any SNP locus, producing four alleles but in practice, SNPs are usually bi-allelic. This is due to the low frequency of substitutions contributing to the low probability that two independent base changes can occur at a single position (Li *et al.* 1981). Additionally, biased-ness of SNP mutations results in either transitions or transversions (Vignal *et al.* 2002). Similarly, SNPs can be found in both regulatory and gene coding regions extending its effect directly or indirectly to the expression of traits. It is noteworthy to know that SNPs in the gene coding regions especially in the exons, can alter the expression of a trait significantly. If the mutations are synonymous/silent mutations, which do not alter the amino acid sequence or inserts an alternative amino acid with similar properties to that of the original amino acid, then the expression of the trait will not be significantly affected. However, there are exceptions to this where the mutation is no longer silent thus, affecting the trait significantly (Goymer 2007). On the other hand, non-synonymous mutations including missense and nonsense mutations, alter the amino acid of a sequence and may either cause alterations in protein structure and function or truncation in the protein, leading to an incomplete and nonfunctional protein. It is the non-synonymous mutations that affect the trait expression in a significant manner (Liu & Cordes 2004).

There are many approaches to the SNP discovery. One of the widely used methods is DNA sequencing due to its simplicity. This methodology is expensive and time consuming especially for discovering large numbers of SNPs and genotyping in large populations. As for SNP genotyping, two main approaches: direct hybridization and two-step separation such as DNA chips, single strand conformation polymorphism (SSCP) analysis and pyrosequencing (Vignal *et al.* 2002) have been undertaken through the years. The type of approaches used for SNP discovery and genotyping is highly dependent on the scale of the project, the type of expertise and budget available. Despite its high developmental cost that limits its application in most aquatic species (Liu & Cordes 2004; Yue 2014), SNP discovery pipelines have been recently developed for some fish species including Atlantic salmon (Hayes *et al.* 2007), gilthead seabream (Cenadelli *et al.* 2007), Atlantic cod (Hubert *et al.* 2010), catfish (Wang *et al.* 2010) and zebrafish (Bradley *et al.* 2007). Over the last decade, microarray-based genotyping technology for SNP markers have been gaining popularity and commercial companies such as Affymetrix and Illumina have developed SNP arrays to conduct high throughput genotyping of SNPs with high accuracy (LaFramboise 2009). Briefly, single-stranded DNA is hybridized to arrays containing hundreds of thousands of unique nucleotide probes and signal intensity associated with each probe and its target are measured. These SNP genotypes are scored using software through allelic discrimination, allowing all existing SNPs in the genome to be analyzed in one experiment (Mao *et al.* 2007). One of the main advantages of using SNP arrays is the ability to rapidly generate large amounts of genotypes due to its simplicity hence, requiring less technical knowledge and resulting in faster processing (Robledo *et al.* 2017). In addition, the bioinformatic analyses are robust and amenable to automation, increasing the reproducibility rate and decreasing the error rate (Hong *et al.* 2012). With the advances of next-generation sequencing (NGS) technologies allowing for discovery of large numbers of SNPs in aquatic species (Yue & Wang 2017), the utility of SNP array in the improvement of traits in aquatic species is definite in the near future despite the high cost of development.

Table 1. Summary of DNA markers, their characteristics, potential applications and limitations

Adapted from (Liu & Cordes 2004)

Marker type	Requires prior molecular information	Mode of inheritance	Type	PCR-based	Genotype scoring methods	Locus under investigation	Likely allele numbers	Polymorphism or power	Major applications	Limitations
Allozyme	Yes	Codominant	Type I	No	Protein electrophoresis (starch gel)	Single	2 to 6	Low	Linkage mapping, population studies	Limited number of loci and alleles per locus Tissue specificity/environmental regulation
Mitochondrial DNA (mtDNA)	No	Maternal inheritance	-	Yes	Gel electrophoresis	-	Multiple haplotypes	-	Maternal lineage	Affected by gender-biased migration and introgression
Restriction fragment length polymorphism (RFLP)	Yes	Codominant	Type I or type II	No; Yes	Southern blotting Gel electrophoresis	Single	2	Low	Linkage mapping	Pre-screen for single copy sequences to be used as probes
Random amplified polymorphic DNA (RAPD)	No	Dominant	Type II	Yes	Gel electrophoresis	Multiple	2	Intermediate	Fingerprinting for population studies, hybrid identification	Sensitive to reaction conditions (reproducibility issues)
Amplified fragment length polymorphism (AFLP)	No	Dominant	Type II	Yes	Gel Electrophoresis	Multiple	2	High	Linkage mapping, population studies	Multi-step, highly technical Radioactive or fluorescent labeling of primers
Microsatellites	Yes	Codominant	Mostly Type II	Yes	Gel electrophoresis	Single	Multiple	High	Linkage mapping, population studies, paternity analysis	Radioactive or fluorescent labeling of primers
Expressed sequence tags (EST)	Yes	Codominant	Type I	Yes	Gel electrophoresis	Single	2	Low	Linkage mapping, physical mapping, comparative mapping	Generation of radiation hybrid panels
Single nucleotide polymorphism (SNP)	Yes	Codominant	Type I or type II	Yes	Gel electrophoresis	Single	2 but up to 4	High	Linkage mapping, population studies	High cost of DNA sequencing
Insertions/deletions (Indels)	Yes	Codominant	Type I or type II	Yes	Gel electrophoresis	Single	2	Low	Linkage mapping	Low number of varying indels in locus of interest

1.2.2.2 Genotype by sequencing (GBS)

With the advent of next-generation sequencing (NGS) technologies, the quality and amount of data generated from the sequencing of genomes have been increased and enhanced immensely. As a result, limitations that bound the utility of traditional marker discovery and genotyping in the breeding of plants and animals have been substantially reduced (Davey *et al.* 2011). This advancement in genomics has allowed for the discovery, sequencing and genotyping of thousands to hundreds of thousands of unbiased markers in large populations even when little or no genetic information is available.

NGS relies on massively parallel sequencing and imaging techniques to yield up to several billions of DNA bases per run (Shendure & Ji 2008) in order to lower the cost of DNA sequencing and make the dream of \$1000 genome closer (Mardis 2006). As a result, several NGS platforms have been developed by companies such as Illumina, Roche, Ion Torrent and Pacific Biosciences. All of these sequencing based marker discovery and genotyping methods generally follow a similar protocol, including DNA template preparation where adapters are ligated to randomly sheared DNA fragments, creating unique DNA barcodes for each fragment. These templates are then pooled and amplified on a synthetic surface to generate billions of sequences in a massively parallel fashion. The incorporation of one or more nucleotides in the sequence is detected via the emission of a signal and these tagged sequences are then separated bio-informatically to its associated samples (Metzker 2010; Davey *et al.* 2011). In this manner, the resources required for the development of markers and the cost of genotyping them per sample will be significantly reduced. Most NGS platforms generate reliable sequences with good coverage, however, the quality of data may differ (Quail *et al.* 2012). For high diversity and larger genome species, portions of the genome can be condensed by reducing its complexity with restriction enzymes and size selection. Some of these methods include reduced-representation libraries (RRLs) and complexity reduction of polymorphic sequences (CRoPS), restriction-site-associated DNA sequencing (RAD-

seq), multiplexed shotgun genotyping (MSG) and genotyping-by-sequencing (GBS) (Davey *et al.* 2011). In particular, GBS has been widely used for generating large numbers of SNPs for use in genetic analyses and genotyping especially in non-model organisms (Figure 1.2) (Elshire *et al.* 2011; Poland & Rife 2012; Beissinger *et al.* 2013).

GBS is a simple, highly multiplexed system to construct RRLs in order to decrease cost, reduce sample handling, decrease PCR and purification steps, remove size fractionation and the limitation of having a reference sequence, and to increase the efficiency of barcoding and the scale of sequencing (Davey *et al.* 2011; Elshire *et al.* 2011). This popularity in utilising GBS has led to expansions of its existing protocol, such as using two restriction enzymes, one common and one rare cutter, *PstI/MspI* to capture more fragments to evenly cover the genome (Poland *et al.* 2012), to altering from sticky end to blunt end ligations in order to bring down the cost of library preparations in maize and rice (Heffelfinger *et al.* 2014) and selective amplification to increase the depth of coverage and number of SNPs discovered in soyabean (Sonah *et al.* 2013). The increasing amount of improvements made in the GBS protocol recently has showed that GBS is becoming a preferable way of producing large amounts of unbiased markers in an inexpensive way, paving the path for the construction of high density and high-resolution genetic maps to enhance population studies such as QTL analysis and genomic selection (Poland & Rife 2012). This increasing interest in using this approach for genomic studies through the discovery and genotyping of sequence-based genetic markers have been increasing in aquatic species as well. They have been used for studies in many fields, including conservation studies for fishes such as rainbow trout (Hohenlohe *et al.* 2013), sturgeons (Ogden *et al.* 2013) and cichlids (Keller *et al.* 2013), genomic analyses in economically important traits such as pigmentation in cichlids (Henning *et al.* 2014), growth in turbot (Wang *et al.* 2015b), large yellow croaker (Xiao *et al.* 2015) and Asian seabass (Wang *et al.* 2015a), sex determination in turbot (Wang *et al.* 2015b), Nile tilapia (Palaikostas *et al.* 2013) and zebrafish (Bradley *et al.* 2011) and disease resistance in channel catfish (Geng *et al.* 2015) and rainbow trout (Palti *et*

al. 2015b). With improvements being frequently made to better the protocol, the utility of GBS for the discovery of genetic markers especially in commercially important species will significantly enhance improvements to MAS or even genomic selection (GS).

Figure 1.2. Principle of genotype by sequencing (GBS)

Example of the genomic regions of two samples containing restriction enzyme cut site (e.g ApeKI) will subsequently generate 'sticky' ends for adaptors (grey) and barcodes (yellow and purple) to be ligated. This produces fragments with barcode+common, barcode+barcode and common+common adaptor combinations. The samples are pooled and only samples sized between 150 - 300bp and contains barcode+common adaptor will be amplified and sequenced. Adapted from (Davey *et al.* 2011)

1.2.3 Construction of linkage maps

A linkage map is a linear arrangement of genes or markers on the chromosome. It is essential in QTL mapping and serves as a genetic map so that the position and distances between these markers on the chromosome are known. This provides us with the means to identify the chromosomal regions or genes that are associated with the desired trait and its relationship with the phenotypic variation in the population through the QTL effect (Doerge 2002; Collard *et al.* 2005). The utility of this linkage map is based on the recombination events and the genetic distance between the markers. This is because markers that are tightly linked or close to each other tend to co-segregate during meiosis and are more likely to be inherited by the offspring. The estimated genetic distance between the markers is calculated by the proportion of recombination events between them. The genetic distance between the markers on a chromosome is expressed in centiMorgan (cM). One centiMorgan means there is a 1% chance of recombination occurring due to crossing over between two markers at one genetic locus on a chromosome in a single generation. The smaller the recombination frequency between two markers, the closer in the genetic distance (Figure 1.3). Markers with recombination frequencies of >50% between them are considered unlinked (Doerge 2002; Collard *et al.* 2005). The conversion from recombination frequencies to cM is carried out using mapping functions such as Kosambi and Haldane as recombination frequency and the frequency of crossing over is not linear related. Kosambi mapping functions assume that crossover event influence adjacent crossover events while Haldane mapping function assumes that there is no interference between recombination events (Kearsey & Pooni 1998; Harker *et al.* 2001).

Figure 1.3. Recombination events between homologous chromosomes during meiosis.

Parental and recombinant (offspring) gametes are denoted as P or R. The distances between the markers E and F and G and H are small and large respectively. Four crossover events are simulated (marked by a grey 'X'). The frequency of recombination increases between markers G and H due to the larger distance between them and this is reflected in the number of recombinants (offspring) in the segregating population. This is in contrast with marker E and F. Reprinted from (Collard *et al.* 2005).

The likelihood of the linkage of markers is determined by the ratios of linkage versus no linkage through statistical testing. The Log 10 ratio of this test, is known as the logarithm of odds (LOD) value or score (Risch 1992). A LOD value above 3 is generally used as a critical value, implying that recombinant ratio of 0.5 (null hypothesis) is rejected and that the possibility of linkage to no linkage is 1000:1. Therefore, LOD value of >3 is deemed as significant linkage and is commonly used as a threshold for the construction of linkage maps (Nyholt 2000).

First and second generation genetic linkage maps have been constructed using conventional marker discovery and genotyping methods for more than 45 aquatic species (Yue 2014) with a particular focus on economically valuable species of fishes. These

included catfish (Waldbieser *et al.* 2001; Liu *et al.* 2003), tilapia (Lee *et al.* 2005; Lee *et al.* 2011; Liu *et al.* 2013a), trout (Gharbi *et al.* 2006; Rexroad *et al.* 2008), carp (Cheng *et al.* 2010; Xia *et al.* 2010), seabream (Franch *et al.* 2006), European seabass (Chistiakov *et al.* 2008), Asian seabass (Wang *et al.* 2011), salmon (Gilbey *et al.* 2004), cod (Moen *et al.* 2009), halibut (Reid *et al.* 2007; Ma *et al.* 2011a), flounder (Coimbra *et al.* 2003; Castaño-Sánchez *et al.* 2010), bivalves including edible oysters (Yu & Guo 2003; Hubert & Hedgecock 2004; Li & Guo 2004; Lallias *et al.* 2007; Shi *et al.* 2009), scallops (Li *et al.* 2005; Qin *et al.* 2007) and abalones (Baranski *et al.* 2006; Sekino & Hara 2007) and crustaceans including shrimps (Wilson *et al.* 2002; Li *et al.* 2003; Zhang *et al.* 2007). Compared to the second generation genetic maps which include more microsatellite markers, single nucleotide polymorphism (SNP) of candidate genes, the first generation genetic maps are of low resolution due to large marker intervals and a low number of markers. Even though the second generation genetic maps may span the genome at higher resolutions due to higher marker number and smaller marker intervals, their application in MAS programs are still limited as conventional methods of marker discovery and genotyping are time-consuming and costly. In order to increase the feasibility of MAS, one of the first approaches is to increase the density of genetic maps. When markers are located close to each other (<5 cM), the possibility of losing a preferred allele due to double recombination events between the markers are very low, hence allowing for transmission of the haplotype through the generations (Rothschild & Ruvinsky 2007). Therefore, increasing efforts are devoted to the construction of high-density genetic maps, encompassing large numbers of genome-wide polymorphic markers by using sequencing-based technologies for marker discovery and genotyping. This allows for a more economically efficient way for increasing the resolution of the genetic maps, as well as the precision in QTL mapping and also enables us to determine the causative mutations and/or location(s) of causative genes (Goddard & Hayes 2009). To date, high density maps were built for several aquatic species such as triangle sail mussel with 4983 markers and average marker interval of 1.8 cM (Bai *et al.* 2016), sea

cucumber with 7839 markers and average marker interval <1 cM (Tian *et al.* 2015b), pacific white shrimp with 6359 marker and average marker interval of 0.7 cM (Yu *et al.* 2015), turbot with 6647 SNPs and average marker interval of 0.4 cM (Wang *et al.* 2015b), Japanese flounder with 12712 SNPs and average marker interval of 0.4 cM (Shao *et al.* 2015), Asian seabass with 3321 SNPs and average marker interval of 0.5 cM (Wang *et al.* 2015a), channel catfish with 29081 SNPs and average marker interval of 0.2 cM (Li *et al.* 2015c), Atlantic salmon with 96396 SNPs and average marker interval of <0.1 cM (Tsai *et al.* 2016), orange-spotted grouper with 4608 SNPs and average marker distance of 0.3 cM (You *et al.* 2013) and Nile tilapia with 3802 SNPs and average marker interval of 0.7 cM. These high-resolution genetic maps provide a valuable tool for fine-mapping QTL with precision as well as for the identification of causative mutations and/or causative genes that are responsible for the economical trait of interest.

1.2.4 Quantitative trait loci (QTL)

Most economic traits such as growth, meat quality, feed conversion ratio and body shape are known as quantitative traits (QT). The phenotypes of this QT vary continuously and they do not show simple Mendelian inheritance in a population that is attributed to a single gene. Additionally, they are controlled by many segregating Mendelian genes all acting together with small to moderate effect and can be influenced by environment and/or gene-environment interactions (Collard *et al.* 2005; Gjedrem & Baranski 2009; Weller 2009). The sections of genomic DNA in the chromosome that contains genes or gene clusters that correlates to this variation in phenotype and controls it, is the quantitative trait loci (QTL) (Geldermann 1975). The identification of the QTL and/or the DNA markers tightly linked to the trait of interest are one of the important factors in contributing to the effectiveness of MAS and the amount of genetic gain. This is especially the case for traits that are difficult to measure such as sex (time required until to reach a certain developmental stage), disease resistance (mortality) and/or are expensive to measure such as meat quality (culling required). There are two ways to

identify QTL: candidate gene approach (Tabor *et al.* 2002) and genome-wide scanning using QTL mapping.

The candidate gene approach relies on prior known information of the genes and its variation that have a sole or large influence on the trait of interest. They are used as anonymous markers but are more targeted to detect variation by the utility of gene sequences of related species for screening (Nielsen *et al.* 2009). This approach has low reproducibility and limited ability to include all possible causative genes of the desired traits especially those polygenic in nature (Tabor *et al.* 2002). Furthermore, the choice of candidate genes is highly subjective considering that the presence whole genome duplication (WGD) of teleost (Van de Peer *et al.* 2009), that may cause genes to rapidly lose or change their functions due to divergence (Glasauer & Neuhauss 2014; Inoue *et al.* 2015). Credible and verified information on well-characterized genes with known functions are required but given the background and complexity of quantitative traits, gene-based approaches greatly limit its application in MAS (Gjedrem & Baranski 2009). The second approach of genome scanning using QTL mapping is the key to overcoming this limitation.

1.2.4.1 QTL mapping

QTL mapping is based on the principle of segregation of markers during crossing over in meiosis. Genes or markers close to each other or tightly linked will be transmitted to the progenies. It does not assume that a known gene or genes are responsible for the variation in the quantitative trait, instead, it is a statistical procedure that correlates phenotypic variation with the chromosomal regions and/or genes, using information from DNA markers that evenly cover the whole genome (Geldermann 1975). Based on the genomic information given from the markers (i.e. genotype), the population is grouped into different genotypic groups. Significant statistical difference between the phenotypic means of the genotypic group ($p < 0.05$) is indicative of the marker being linked to the QTL controlling the trait (Tanksley 1993; Young 1996). The significance of the statistics, p -value indicates that the linkage of the marker to the QTL is due to

recombination. This is because as the marker distance from the QTL decreases, the chances of recombination are lowered and the QTL and marker will be inherited together. A loosely linked or unlinked marker will not have statistically significant means ($p > 0.05$) as they are randomly inherited because they are far away from the QTL or on a different chromosome (Figure 1.4). The position and size of the QTL can be identified through joint analysis of two or more markers and a genetic map.

Figure 1.4. The principle of QTL mapping.

Diagram indicating tight and loose linkage of the markers to the QTL. (a) Recombination events are indicated by a grey dotted cross. (b) Gametes in the population. (Top to bottom) Marker E is tightly linked to QTL and tends to inherit together with QTL. In contrast, Marker H is loosely linked to QTL and tends to be inherited randomly. Thus, the marker (i.e. Marker E) linked to

QTL or a gene controlling a particular trait will present significant differences in the partitioned genotypic groups based on the genotype of the marker. The linkage of Marker E to a QTL showed a significant difference between the means when compared to marker H, where no significance was observed. Adapted from (Collard *et al.* 2005).

1.2.4.1.1 Methods of analysis of effects and locations of QTL

The detection of QTL is performed through several statistical methods to search systematically for QTLs through analyzing the mapping data and phenotypic data. Three of the most widely used methods are single-marker analysis, simple interval mapping (SIM) and composite interval mapping (CIM) (Tanksley 1993; Liu 1997). Amongst these three, SIM (Lander & Botstein 1989) is one of the current standard mapping used for mapping QTLs. It analyzes the intervals between neighbouring pairs of linked markers across the chromosomes of the genetic map in an incremental manner to test for the absence and presence of a putative QTL. Using linked markers for analysis compensates for recombinations between markers and the QTL, making it more statistically powerful than single-marker analysis where the chances of the QTL being detected and the effect of the QTL decreases with increasing marker distance. The SIM analysis is based on likelihood ratios and predicts the possible sites for QTL between the neighbouring linked markers on a genetic map. A dense grid of possible QTL sites, calculated from genotypes with respect to marker genotyping data of an individual is generated. The calculated likelihood ratio is presented as the logarithm of odds (LOD) scores that are used to determine the most likely position of the QTL in the genetic map (Collard *et al.* 2005). Despite being one of the widely used methods for detecting QTL, the SIM approach can be biased and QTLs can be mapped to wrong positions especially when two or more QTLs are located on the chromosome (Haley & Knott 1992; Martínez & Curnow 1992). The utility of SIM was superseded by CIM for cases where multiple QTLs and/or linked QTLs were suspected such as for polygenic traits. CIM is a combination of interval mapping with linear regression and adjacent linked markers and markers significantly associated with the trait are used as cofactors (Jansen 1993; Zeng

1993; Jansen & Stam 1994; Zeng 1994). This allows for some control for genetic variation in other regions of the genome and reduces the background "noise" that may affect QTL detection hence, it is more precise and effective at mapping QTL (Collard *et al.* 2005). A combinatorial method to improve the accuracy of QTL detection could be achieved by firstly performing initial genome scans using SIM approach as it has the potential to indicate multiple or "ghost" QTL and markers linked to the trait, then utilizing the identified markers as cofactors for detecting QTL using CIM approach (Doerge 2002) (Figure 1.5). Regardless of the type of computational analysis used for the detection and localization of QTL, the significance threshold in which the QTL is determined to be statistically significant is mostly performed using permutation tests (Churchill & Doerge 1994). The permutation test is a robust testing to determine the test statistics for the mapping experiment. It randomly 'shuffles' the phenotypic values of the population and keeps the marker genotypic values constant, resulting in a broken association between the marker and the trait. If the association between the marker and trait is broken, the original association is deemed to be real. If the association remains, it reflects more of a random association between marker and trait and is not indicative of a 'real' QTL (Doerge & Rebai 1996). The process is repeated a minimum of 1000 times to assess the levels of false positive marker-trait association in the QTL analysis in order to determine the significance level (95% significance level or $p < 0.05$). These test statistics (i.e. calculated LOD scores) generated from permutation test are compared to the LOD scores from QTL mapping analyses to identify the QTL. LOD scores from mapping analysis that surpasses the calculated LOD score of the significance level of 95% are deemed as real QTL (Figure 1.5) (Doerge 2002; Collard *et al.* 2005).

Figure 1.5. Three of the widely used QTL mapping analysis.

QTL mapping was carried out using three methods; Single marker analysis, SIM, and CIM. The single marker analysis identified one point/one marker (black arrow), the SIM identified four QTL (blue arrow) and CIM identified two QTL (green arrow) on the LOD profile. The red line represents the calculated LOD score using permutation testing (i.e. significance of 95%). Adapted from (Doerge 2002).

In general, QTL may be classified into 3 categories: suggestive, significant and highly significant where significance levels vary from once at a random mapping study to significance levels of 5 to 0.1%, respectively (Lander & Kruglyak 1995; Collard *et al.* 2005). The detection of QTL in segregating populations can be affected by many factors such as the magnitude of the QTL effect, environmental effects, population size and experimental error, including errors in marker genotyping and phenotypic evaluation (Collard *et al.* 2005). One of the most important factors is the population size which is indicative of the statistical power, estimation of gene effects and confidence intervals of QTL locations (Beavis 1998). The aim of QTL mapping is to provide a framework of genetic information across the genome for the identification of the markers linked to the trait of interest. Therefore, the resolution of the QTL mapping is crucial, especially for the application of MAS. Typically, in fishes, family-based QTL mapping is carried out. As the detection of the QTL is based on recombination events between the marker and the QTL, which can vary between families and generations, the reliability and usefulness

of the marker linked to the QTL may be compromised. This limitation is one of the greatest strength and also the greatest weakness in family-based QTL mapping. However, enlarging the population size and increasing the density of markers on the genetic map so that more tightly linked markers could be discovered, may overcome this limitation. Carrying out high-resolution mapping or fine-mapping of QTL decreases the confidence interval of the QTL and increases the chance of localizing a causative mutation or gene that is associated with the trait of interest thus, allowing for a better understanding of the quantitative trait of interest (Collard *et al.* 2005).

1.2.4.2 Gene-based/Candidate gene approach in mapping traits

The rapid development of modern genetics especially in gene expression studies has allowed scientists to understand complex quantitative traits and positively impact in stock quality and productivity through modifications in these traits (McAndrew & Napier 2011). In this respect, if the trait of interest has been shown to have a genetic basis albeit partially, it can be explored at the molecular level. This approach allows for discovery of important genes and polymorphisms that may influence individual variations in the trait of interest. Global gene expression patterns under a range of different conditions may identify patterns correlating differences in traits of interest such as growth, disease resistance and lipid composition of fishes (Haley & De Koning 2006). The recent advances in sequencing technology such as RNA sequencing, which will be discussed in the next section, significantly improve our knowledge on the gene regulation in the transcriptome of the fishes under different conditions. Additionally, other modern sequencing techniques such as genome sequencing enrich the transcriptomic data and can significantly impact stock quality and productivity. Currently, genome sequencing has been already been conducted in several farmed aquatic species including rainbow trout (Berthelot *et al.* 2014), Nile tilapia (Brawand *et al.* 2014; Conte *et al.* 2017), channel catfish (Liu *et al.* 2016), Atlantic salmon (Davidson *et al.* 2010), Atlantic cod (Star *et al.* 2011), Asian seabass (Vij *et al.* 2016), European

seabass (Tine *et al.* 2014), turbot (Figueras *et al.* 2016) and common carp (Xu *et al.* 2014b) and genome sequencing and assembly for more than 20 aquaculture species are underway (Liu *et al.* 2012b; Yue & Wang 2017).

1.2.4.2.1 Transcriptomic analysis of aquatic species: RNA sequencing

The transcriptome is a complete and dynamic set of transcripts that vary depending on the specific developmental stage or physiological condition of the organism and can be influenced by environmental effects (Wang *et al.* 2009). These transcripts include protein-coding messenger RNA (mRNA) and non-coding RNA (ncRNA) including ribosomal RNA (rRNA), transfer RNA (tRNA) and other ncRNAs (Lindberg & Lundeberg 2010). The analysis of the transcriptome is a powerful tool for the dissection of the relationship between genotype and phenotype and allows us to understand the underlying mechanisms and pathways that control the variation of the phenotype in the population. In the study of transcriptomics, it is not limited to only the quantification of gene expression levels amongst samples but also the mapping and annotation of the transcriptome and the functional determination of each gene (Wang *et al.* 2009; Costa *et al.* 2010). Of the many approaches developed for large-scale transcriptome profiling including hybridization or sequence based, RNA sequencing (RNA-Seq), one of the next generation platforms for massively parallel sequencing is the most popular due to its declining cost and advantages over hybridization-based approaches (Wang *et al.* 2009; Wilhelm & Landry 2009). In comparison to hybridization-based technologies such as microarray or tiling arrays, that function by hybridizing labeled DNA or RNA target molecules to probes that are fixed onto a solid surface, sequence-based approaches directly determine and quantify the complementary DNA (cDNA) generated from mRNA. Even though hybridisation based approaches are relatively high throughput and inexpensive, its applications are limited especially in non-model fishes, due to several factors including dependence on existing knowledge of genomic sequence, high background noise due to nonspecific hybridization, signal saturation for certain

transcripts and difficulty in conducting comparative studies in gene expression levels across different experiments (Marioni *et al.* 2008; Wang *et al.* 2009; Bradford *et al.* 2010). Sequencing-based approaches previously utilising Sanger sequencing and tag-based methods such as serial analysis of gene expression (SAGE) (Matsumura *et al.* 2005) or massively parallel signature sequencing (MPSS) (Brenner *et al.* 2000) have been rapidly replaced by RNA-seq due to its higher throughput, lower cost and the potential to capture nearly complete information of the transcriptome dynamics including different isoforms (Wang *et al.* 2009).

RNA-seq employs next-generation sequencing (NGS) technologies to conduct massive parallel sequencing of cDNA from an RNA sample. Transcriptomic analyses using RNA-seq involves three steps including library construction, sequencing on specific NGS platform and bioinformatic analysis (Figure 1.6). The library construction is one of the key steps as it determines the accuracy of which the final sequencing data reflects the original transcriptome. The RNA of the samples are extracted and rRNA are removed for the enrichment of mRNA molecules, either through oligo d(T) purification or rRNA depletion by probe hybridization or exonuclease activity. The clean RNA is fragmented, converted into cDNA libraries, amplified and sequenced from either a single end (SE) or paired-end (PE) on different sequencing platforms. A typical read length of 30-400 bp which is dependent on the type of platform used, will be obtained and these sequences known as reads, will be aligned either to a reference genome or transcriptome or assembled *de novo* in the absence of an available reference genome using bioinformatic tools. The expression levels, structure and functional annotation of the transcripts can then be determined. (Wang *et al.* 2009; Costa *et al.* 2010). There are several crucial factors to consider when designing and carrying out RNA-Seq studies: 1) Gene expression is limited to the point of RNA extraction hence, the type and number of tissues and the time point or developmental stage must be appropriate in order to reflect the biological process in question. 2) Biological replicates are important both statistically and for assigning confidence to expression levels since different protocols

in generating the library may introduce bias in the final data obtained (Wang *et al.* 2009).

3) Mean sequencing coverage is typically 30 times, meaning 95% of the genome being covered will have a minimum of 10 times coverage. However, the coverage is dependent on the heterozygosity of the transcriptome including rare and lowly expressed transcripts, where higher sequencing coverage are required as the transcriptome complexity increases (Tarazona *et al.* 2011; Sims *et al.* 2014). 4) Read lengths and the type of reads (SE or PE) is dependent on the organism, the biological question in study, the cost involved and the error rate. Generally, studies investigating gene expression levels, especially of known genes of well-annotated genome assemblies require short read lengths and SE sequencing at lower coverage hence, are generally lower in cost to conduct. In contrast, species without reference genome or annotated genome require longer read lengths, PE sequencing with a higher coverage than 30 times as there is little knowledge of the gene structure hence, increasing the cost concurrently (Wang *et al.* 2009; Costa *et al.* 2010; Oszolak & Milos 2011).

Figure 1.6. A simplified overview of RNA-seq and data analysis.

A. RNA isolation is first carried out with the biological material of choice. Specific RNA molecules are then isolated using different protocols such as poly-A enrichment for polyadenylated transcripts. Next, reverse transcription is carried out to convert RNA to complementary DNA (cDNA) and ligation of sequencing adaptors are performed on the ends of the cDNA fragments. Lastly, the ligated cDNA fragments are PCR amplified and the RNA-Seq library is constructed and ready for sequencing. B. The sequencing reads are typically aligned to a reference genome. This may be carried out through the assembling of transcripts using reference annotations or *de novo* assembly approaches. Following that, the expression level of each gene is estimated by counting the number of reads that align to each exon or full-length transcript. Further downstream analyses with RNA-Seq data include testing for differential expression between samples, detecting allele-specific expression, and identifying expression quantitative trait loci (eQTLs). Reprinted from (Kukurba & Montgomery 2015)

In fishes, RNA-Seq have been quickly adopted for various studies including, evolution (Ma *et al.* 2016b; Pasquier *et al.* 2016), mating tactics (Schunter *et al.* 2014),

effects of environmental stress such as temperature (Liu *et al.* 2013b; Smith *et al.* 2013), salinity changes (Xu *et al.* 2015; Nguyen *et al.* 2016), hypoxia (Liao *et al.* 2013; Zhang *et al.* 2016) and economically important traits such as growth (Robledo *et al.* 2016), skin color (Zhu *et al.* 2016), disease susceptibility (Xia *et al.* 2013), lipid composition (Hu *et al.* 2016a) and sex determination (Böhne *et al.* 2014). With the continuing efforts to sequence 1000 species under the Fish-T1k project (Sun *et al.* 2016) and improving sequencing technologies, our knowledge in the relationship between genotype and phenotype will accelerate MAS and improve stock quality and productivity significantly.

1.3 Growth of fishes and the candidate genes

Somatic growth is a complex trait and one of the important and desired traits in aquaculture and its heritability adds to the complexity (Reed *et al.* 2015). Traditional selections that are based on phenotype for growth are effective but they are time-consuming, labour-intensive and can be influenced by environmental factors (Gjedrem & Baranski 2009). In aquaculture, the growth of the fish is measured in terms of weight and length but the main value of understanding growth lies in its most valuable and edible part, the flesh or better known as the muscle. Therefore, an insight into the genetic basis of growth and its regulation in fishes will potentially be beneficial in the selection and reap significant benefits in the productivity, sustainability as well as profitability.

1.3.1 Factors involved in growth regulation of fishes

Fishes have indeterminate growth patterns with increasing body size and muscle mass till mortality. Growth is under the influence of several physiological pathways that are sensitive to environmental cues, nutritional status and involved in energy and metabolism (De-Santis & Jerry 2007). Somatotrophic axis, involving the growth hormone (GH) and insulin growth factors (IGF) and its associated factors plays a major role in the regulation of growth of fishes (Canosa *et al.* 2007). The GH-IGF system in fishes is

complex with many subtypes of receptors and binding proteins due to genome duplication events that occurred during the evolution of teleost (Jiao *et al.* 2006; Hagemeister & Sheridan 2008; Roch *et al.* 2009). The factors involved in the GH-IGF system are widely distributed and regulate both growth and other biological processes such as protein metabolism, osmoregulation, reproduction, behaviour, immunity and lipid metabolism (De-Santis & Jerry 2007; Duan *et al.* 2010; McMenamin *et al.* 2013). The complexity of this regulation is further intensified as some components appear to be organism and tissue-specific (Canosa *et al.* 2007) and dependent on genetic background (Devlin *et al.* 2001; Biga & Meyer 2009). One of the major factors under the stimulatory control of the growth hormone, IGF appear to be associated with growth and metabolism (Castillo *et al.* 2004; Codina *et al.* 2008) via the mitogen-activated protein kinase (MAPK) and phosphatidylinositol 3-kinase (PI3 kinase) signalling pathways (Pozios *et al.* 2001). Studies in commercial food fishes such as Nile tilapia (Huang *et al.* 2012), smooth tongue sole (Ma *et al.* 2011b) and channel catfish (Peterson *et al.* 2004) have observed an up-regulation in these somatotropic genes. In order to identify key genes underlying the growth process in fishes, studies were also focused on gene expression where growth was reduced in fishes under different conditions such as starvation (Rescan *et al.* 2007), growth hormone transgenesis and domestication (Devlin *et al.* 2013), compensatory growth (Ali *et al.* 2003; Kamei *et al.* 2011) and sex (Manor *et al.* 2015). This suggests the complexity of growth and its interaction with other biological pathways that are involved in normal and rapid growth. Therefore, a full understanding of GH-IGF function in a variety of physiological processes in different species such as growth, metabolism, reproduction and etc will require more work due to the different forms of its associated factors. The diversity that is associated with the evolution of teleost will undoubtedly make them useful targets for further exploration due to their economic importance in aquaculture, especially in selection procedures during breeding.

1.3.2 Myotomal growth in fishes

The main value of fish in aquaculture is the myotome or skeletal muscle which comprises a great proportion (40-60%) of the mass of the fish. It is composed of fast-twitch (white) and slow-twitch (red) fibers (Valente *et al.* 2013) that are critical for survival as it is involved in foraging, escaping from prey and migration (Videler 2011). Muscle growth or myogenesis in many fish species occurs at an early stage of development through myogenic regulatory factors (MRFs) such as *myoD*, *mrf4*, *mrf5*, myogenin (Hinits & Hughes 2007; Hinits *et al.* 2009). The composition of muscles is primarily muscle fibers and the process of post-embryonic muscle cell differentiation into fast and slow muscle fiber is dependent on the activity of hedgehog (Hh) proteins (Wolff *et al.* 2003) and migration (Henry & Amacher 2004). In adult fishes, hyperplasia (recruitment of new muscle fibers) or hypertrophy (increase in size of muscle fibers) contributes to the ultimate body size of the fish (Weatherley *et al.* 1988; Rowleson & Veggetti 2001). This recruitment and/or increase in the size of muscle fibers is highly related to metabolism and the growth of fishes and can be easily affected by environmental factors (Johnston 1999). Examples of myogenic factors involved in the metabolism and growth of fishes are *myoD* and TGF- β molecules including myostatin (Bellingue *et al.* 2005) and activin (Lee *et al.* 2012). Developmental genes such as *pax 3* and *pax 7* (Kacperczyk *et al.* 2009; Kacperczyk *et al.* 2011) and genes that control protein turnovers such as F-box proteins, caspase, and calpain (Cleveland & Weber 2013) also play a role in hypertrophic and hyperplastic growth. One signalling pathway, the Hippo pathway, first characterised in *Drosophila melanogaster* (Harvey & Tapon 2007), has been associated with maintenance of skeletal muscle mass and fibre (Watt *et al.* 2015) and the body shape of vertebrates (Porazinski *et al.* 2015) suggesting its importance in the development of somatic growth of skeletal muscle in fishes. Being able to understand the molecular mechanisms governing the development of muscle fibers may allow us to further explore the ways to harness this knowledge in increasing the muscle mass of the fishes for production in aquaculture.

1.3.3 Sexual dimorphism and growth

Sexual size dimorphism is common and varies between species in fishes (Fairbairn 1997; Badyaev 2002). Fishes exhibit sexual dimorphism through variance in size, colour, body proportions and they are generally linked to reproduction. Additionally, other factors such as sexual selection, dimorphic niches, and food competition may also play a role in sexual size dimorphism (Hedrick & Temeles 1989; Blanckenhorn 2005). In teleost, the size differences can be very substantial either biased towards bigger females such as in flounder (Kobelkowsky 2004), sole (Shao *et al.* 2014), Atlantic halibut (Hjalti-Jákupsstovu & Haug 1988) or bigger males such as in cichlids (Schütz & Taborsky 2000) including tilapia (Bhatta *et al.* 2012). Studies have also shown a correlation between growth and sex in eels (Degani *et al.* 2003) and tilapia (Davis *et al.* 2008) as well as a correlation between sex maturity and growth in sole (Ji *et al.* 2011), Atlantic halibut (Hagen *et al.* 2006) and Atlantic salmon (Leclercq *et al.* 2010). The understanding of the molecular mechanisms of sexual size dimorphism is complex and still at its infancy, however, the advantages of having better growth performance and bigger body size has pushed for the culture of the sex that marks superior growth. Monosex cultures in aquaculture have been practised for a variety of species due to improved growth performance; all male culture such as tilapia (Beardmore *et al.* 2001) and channel catfish (Goudie *et al.* 1994) or all female culture such as trout (Sheehan *et al.* 1999) and carp (Cherfas *et al.* 1996). Besides the economic benefits of better growth performance, monosex cultures prevent uncontrolled reproduction, undesirable impacts on flesh quality and aggressive interactions (Beardmore *et al.* 2001). Improved understanding of sexual size dimorphism would be useful in determining the genetic impact of such monosex cultures in the management of cultured fishes.

1.3.4 Status of mapping QTL and transcriptome analysis of growth in aquaculture species

Growth and its related traits have played an important role in the economics in aquaculture and have prompted many quantitative and transcriptomic studies of their genetic basis in order to successfully apply them in breeding programmes. It still remains one of the most popular traits in study even though early work on selection of increased growth rate through mass crosses in common carp (Moav & Wohlfarth 1976) and Nile tilapia (Hulata *et al.* 1986) have not been successful due to the low genetic variation caused by inbreeding (Bentsen & Olesen 2002). Major and minor QTL for growth traits have been detected in many species throughout the years with varying phenotypic effects on the populations in study (Yue 2014). In Asian seabass, one major and 5 minor QTL were detected from a high density genetic map consisting of 3321 SNPs. The major QTL was detected on linkage group (LG) 5, explained 16% of the total phenotypic variation and a candidate gene, *acox1* was identified (Wang *et al.* 2015a). In rainbow trout, QTL for growth-related traits have been conducted and verified on 10 different LG and two major growth related candidate genes, *pax7* and *gh2* have been identified (O'Malley *et al.* 2003; Wringe *et al.* 2010). Further studies combining microarray and QTL analysis have also uncovered tissue-specific and size-related genes (Kocmarek *et al.* 2015). In common carp, four major QTLs for growth-related traits explained >20% phenotypic variation (Laghari *et al.* 2015) and studies using eight distinct families showed that QTL for growth traits detected can be common in multiple families (Lv *et al.* 2016). The increasing genomic resources available had allowed scientist to be able to compare data across species and increase the density of genetic maps. In turbot, previous studies on QTL for growth traits have not been significant due to a difference in family structure, analysis methodology and low resolution linkage maps (Ruan *et al.* 2010; Sánchez-Molano *et al.* 2011). Recent studies enhanced by the availability of NGS technology, increased the confidence level of detected QTL for growth traits significantly with LOD

values of >5 for major QTL and identified the SNP markers within these QTL (Wang *et al.* 2015b). In flounder, a high density genetic map consisting of 1478 microsatellites, identified four QTL for growth traits explaining more than 13% phenotypic variation on LG 14 (Song *et al.* 2012). In Atlantic salmon, different mating models and families and newer marker technology such as SNP chip detected more QTL for growth (Baranski *et al.* 2010; Gutierrez *et al.* 2012) when compared to previous studies which, only identified one QTL for body weight in two LGs (AS 8 and AS 11) (Reid *et al.* 2004). The underlying cause of the QTL effect are inferred through comparative syntenic analyses with well-characterized candidate gene, model species or closely related species. However, QTL spans large sections of chromosome and can contain many candidate genes hence, some studies have chosen instead to use the candidate gene approach to identify molecular markers or growth-related genes. Generally, candidate genes have been focused on somatotrophic axis as a regulator of growth. Polymorphisms in the genes regulating the axis such as GH, IGF, myostatin (MSTN) have been reported to be associated to growth in several fishes such as Arctic charr (Tao & Boulding 2003), Atlantic salmon (Gross & Nilsson 1999; Tsai *et al.* 2014), largemouth bass (Li *et al.* 2009) and common and bighead carp (Liu *et al.* 2012a; Sun *et al.* 2012). The advancement of sequencing technology such as RNA-Seq has allowed the association of phenotype and genotype to become more straightforward. Using RNA-Seq technology, 22 and 45 SNP markers associated with growth traits were reported in rainbow trout (Salem *et al.* 2012) and turbot (Robledo *et al.* 2016) respectively. RNA-Seq was also utilized to identify genome-wide gene-associated SNP markers in catfish to generate a SNP chip to increase the efficiency of large-scale genotyping for selection programmes for many economical traits including growth (Liu *et al.* 2011; Liu *et al.* 2014c). The studies on QTL mapping for growth has wide-reaching implications as growth rate affects survival and fecundity across many taxa (Arendt 1997) including avoidance of predation (Sundström *et al.* 2005) and utility of ephemeral resources (Biro *et al.* 2006). Furthermore, growth and stocking density is of importance to commercial

aquaculture especially those utilising re-circulating systems and the ability to culture fast growing fishes rapidly maximises its production, profitability and sustainability (Merino *et al.* 2007).

1.4 Meat quality: Lipid content of fishes

Lipids especially fatty acids (FA) are crucial components in the diet of animals including fish and humans (Holman 1998; Sargent *et al.* 1999). Despite the different types of FA, concerted efforts have been placed on the essential fatty acid families, omega 3 (n-3) and omega 6 (n-6) polyunsaturated fatty acids (PUFA) due to their ability to assist in management and prevention of chronic diseases (Simopoulos 2006; Serhan *et al.* 2008; Wall *et al.* 2010), especially for cardiovascular diseases (Holub & Holub 2004; Biscione *et al.* 2007). Fish meat is an important resource of a human's diet due to its high quality of animal protein, essential lipids as well as vitamins, minerals and trace elements (Sargent & Tacon 1999; Arts *et al.* 2001; Kris-Etherton *et al.* 2002). Additionally, the lipid content in the muscles of the fish is one of the key determinants to enhancing the eating quality of the fish (Robb *et al.* 2002; Lefevre *et al.* 2015). The demand for healthy fishes with high lipid content especially polyunsaturated fatty acids (PUFA) and good eating quality makes it an important economical trait in aquaculture.

1.4.1 Total Lipid content in fishes

Generally, lipids can be divided into two major groups: the membrane lipids and the triacylglycerols (Figure 1.7).

Figure 1.7. Most common types of storage and membrane lipids.

The two major groups of lipids are the storage and membrane lipids. Triacylglycerides make up the bulk of the storage lipids (purple) while membrane lipids are further categorized into phospholipids and glycolipids (blue). The phospholipids and glycolipids are further subcategorised into glycerolphospholipids and sphingolipids (purple) and sphingolipids and galactolipids respectively (purple). Adapted from (Nelson *et al.* 2010)

The membrane lipids are an integral component of cell membranes while the triacylglycerides are energy reserves. Fatty acids make up the bulk of the triacylglycerides. Similarly to mammals, in fishes, lipids are synthesized in the liver and transported to peripheral tissues such as muscle (Sheridan 1988). Several metabolic processes such as lipogenesis, lipid transportation via lipoproteins and lipid storage regulate the lipid deposition in different tissues in the fish (Sheridan 1988; Tocher 2003). Lipids together with its associated fatty acids are one of the main sources of metabolic energy for growth, reproduction and are vital for health and immunity of the fish (Sargent *et al.* 2003; Tocher 2003; Arts & Kohler 2009). Previous studies have also shown that the concentration of lipid varies considerably in the different parts of the body (Mráz & Pickova 2008) and they are preferred energy source for growth and energy due to the protein sparing effect (Kaushik 1998; Sargent & Tacon 1999; Kolditz *et al.* 2008). Variations in the lipid composition of fishes can be observed due to the environment, nutrition, strain, sexual maturity, and age (Vergara *et al.* 1996; Kiessling *et al.* 2001; Grigorakis *et al.* 2002; Olsson *et al.* 2003). In particular, diet changes which affect the nutrition of the fish causes changes in fat deposition and growth performance

(Yamamoto *et al.* 2002). Besides the extrinsic factors that may affect the lipid content of the fish, studies in fishes such as Atlantic salmon (Leaver *et al.* 2011) and rainbow trout (Tobin *et al.* 2006) have observed that total lipids are heritable, implicating that differences in total lipids are genetic in origin. These differences seen in individual fishes and heritability of total lipids necessitate further probing to enhance selection procedures for healthy fishes with higher lipid content and good growth performance.

1.4.2 Biosynthesis and metabolism of fatty acids

The most important aspect of fish nutrition that is of particular interest to consumers is the fatty acid composition in the fishes, particularly in the muscle. Fatty acids in fish can arise mainly from dietary sources but it can be synthesized from within the fish through lipogenesis as well. The main tissue where lipogenesis takes place is in the liver (Henderson & Tocher 1987) and fishes modulate the expression of lipid metabolic genes during dietary lipid intake (Alves Martins *et al.* 2012). The actual rate of fatty acid synthesis endogenously is inversely related to the level of lipid in the diet, and marine fishes that are provided with a high lipid diet have less *de novo* lipogenesis compared to freshwater fish (Tocher 2003).

Similarly to mammals, lipogenesis in fishes begins with the interaction of factors from metabolic pathways such as pentose phosphate pathway (PPP) and tricarboxylic acid cycle (TCA) and multi-complex of fatty acid synthases (FAS). This results in saturated fatty acids (SFA), C₁₆ (palmitic acid, 16:0) and C₁₈ (Stearic acid C18:0) which are then converted with the help of delta 9 (Δ 9) stearoyl-CoA desaturase (SCD) in the endoplasmic reticulum, to monounsaturated fatty acids (MUFA) such as C18:1n-7 or C16:1n-7 (Sargent *et al.* 2003; Tocher & Glencross 2015). These SFA and MUFA can be elongated to longer chain SFA and MUFA through the action of elongases (Elovl, elongation of very long fatty acids). However, all vertebrates lack delta 12 (Δ 12) and delta 15 (Δ 15) desaturases, therefore, fishes are not able to synthesize polyunsaturated fatty acids (PUFA) endogenously (Tocher 2003) and hence require

them in their diet. The dietary precursor C18 PUFA's are then desaturated through delta 4 to delta 6 fatty acyl desaturase ($\Delta 4$ - $\Delta 6$ fads) and elongated through elongases (Elovl2, 4, 5) into long chained C20-C22 PUFA's (Figure 1.8) (Castro *et al.* 2012). To date, elongases have been cloned from many freshwater species including zebrafish, common carp, tilapia, and marine species including Atlantic cod, turbot, Atlantic salmon and gilthead seabream (Agaba *et al.* 2004; Hastings *et al.* 2004; Agaba *et al.* 2005).

Figure 1.8. Biosynthetic pathway of very long chained polyunsaturated fatty acids (VLC-PUFA).

A condensed version of the pathways of long-chain polyunsaturated fatty acid biosynthesis in fish with more focus on the production of omega 3 and 6. Enzymes are referred to as either 'ELOVL' for fatty acid elongase (e.g. ELOVL5 is $\Delta 5$ fatty acid elongase) or 'FADS' for fatty acid desaturase (e.g. $\Delta 5$ FADS is $\Delta 5$ fatty acid desaturase). Names and abbreviations for the omega 3 and 6 fatty acids and their precursors are ALA: 18:3n3, EPA: 20:5n3 and DHA: 22:6n3 and LA; 18:2n6 and ARA: 20:4n6 respectively. Fatty acids and the precursors important to human health are abbreviated and denoted in red and blue respectively (Castro *et al.* 2012).

Two particular points to note in the synthesis of DHA from EPA: firstly, it occurs via 2 pathways, the "Sprecher's shunt" involving the action of delta 6 ($\Delta 6$)

desaturase (Sprecher 2000) reported in trout (Buzzi *et al.* 1997) and the recently identified delta 4 ($\Delta 4$) desaturase pathway in rabbit fish (Li *et al.* 2010). Secondly, delta 8 ($\Delta 8$) desaturase had activities characteristic of $\Delta 6$ desaturase (Monroig *et al.* 2011) in both freshwater and marine fishes. Lastly, functional delta 5 ($\Delta 5$) desaturases are reported to be present in freshwater fishes (Nakamura & Nara 2004), but absent in marine fishes such as seabass, with the exception of rabbit fish (Li *et al.* 2010) hence, marine fishes are fully dependent on the dietary inputs of PUFA for their nutrition (Tocher & Glencross 2015). It is generally accepted that the highly diverse $\Delta 6$ desaturation (Castro *et al.* 2012) has been regarded as a rate-limiting step (Brenner 1981) in the synthesis of PUFA, the health of the fishes and the economic value of understanding the biosynthesis of PUFA had prompted the characterization of fatty acid desaturase 2 (Fads2) enzymes with $\Delta 6$ desaturases. To date, more than ten Fads2 enzymes with $\Delta 6$ desaturases from freshwater and marine fishes such as common carp (Ren *et al.* 2012), Nile tilapia (Tanomman *et al.* 2013), Asian seabass (Mohd-Yusof *et al.* 2010), Atlantic bluefin tuna (Morais *et al.* 2011) and Atlantic cod (Tocher *et al.* 2006) had been identified. Additionally, Fads2 with monofunctional $\Delta 5$ desaturase activity and bifunctional $\Delta 6/\Delta 5$ desaturase activity were reported in Atlantic salmon (Hastings *et al.* 2004) and zebrafish (Hastings *et al.* 2001). This diversity in activity is thought to be linked to food sources in the habitat (Castro *et al.* 2012). The biosynthesis of PUFA in fishes has become an important topic in aquaculture due to being the main source of omega 3 fatty acids (i.e. EPA and DHA) in human and other animal diets (Simopoulos 1999). Additionally, the cost of feeding marine fishes with high PUFA content species such as anchovies and capelins are increasing with the dwindling wild supplies (Sargent & Tacon 1999). Therefore, increasing the knowledge of the molecular genetic basis underlying the phenotypic variation in PUFA biosynthesis will allow us to make well-informed decisions when it comes to the selection of fishes in breeding programs.

1.4.3 Mapping QTL and transcriptome analysis for lipid composition in aquaculture species

Lipid or fat composition, including total lipid content, fatty acid content have been one of the more challenging topics in aquaculture and progress have been slow in this field in regards to quantitative and transcriptomic studies. This is mainly because accurate measurements of lipid-related traits can only be conducted in slaughtered individuals and the high cost of chemical analysis, hence, mostly restricting to selection within the family, limiting its utility in breeding programmes (Derayat *et al.* 2007; Baranski *et al.* 2010; Hamzah *et al.* 2016). Even though aquatic species including fish and shellfish contain less than 2.5% total lipid, lipid content has been reported to be heritable, mostly moderately in fishes including rainbow trout (Gjerde & Schaeffer 1989; Tobin *et al.* 2006), Atlantic salmon (Leaver *et al.* 2011), European whitefish (Kause *et al.* 2011), European seabass (Saillant *et al.* 2009), Nile tilapia (Hamzah *et al.* 2016) and common carp (Kocour *et al.* 2007). This heritability is thus indicative of the possibility of successful selection in breeding programmes.

In Asian seabass, one major and fifty-three minor QTL for lipid composition and water content have been detected (Xia *et al.* 2014a). One of the candidate genes identified from the major QTL located on LG 5 was intestinal fatty acid binding protein a (*ifabp-a*), which was reported to be involved in fatty acid transport. Further analysis identified the association of *ifabp-a* and its isoform, *ifabp-b* with growth in the Asian seabass. The overlapping of QTL intervals and its association with growth traits suggested that the lipid content and composition were determined by several chromosomal regions with relatively small effect (Xia *et al.* 2014a; Yue 2014). In common carp, one major and three minor QTL explaining 36.2% and 15.3-19.5% phenotypic variance was detected in four different LG. Even though different family structures were utilized in this study, recurring QTL have been identified in these analyses, increasing the credibility of the location of these QTL (Kuang *et al.* 2015). In

Atlantic salmon, major QTL for fat percentage was detected on LG16 and a marker, Ssa0016NVH was identified to be tightly linked to it despite the absence of mentions on effect on the phenotypic variations of the population studied (Derayat *et al.* 2007; Yue 2014). This is however not in agreement with recent studies of the Atlantic salmon using SNP arrays, where four and one SNP on LG 10 and 9, respectively, were identified to be significantly associated with the variation in fillet fat content (Sodeland *et al.* 2013). Nonetheless, the data in both studies are not comparable as QTL mapping studies are limited to genetic diversity provided by parents in a segregating population and are lower in resolution compared to genome-wide association studies (GWAS), where hundreds of thousands of markers are utilized to detect their association with the trait (Bush & Moore 2012). Lipid deposition in fishes is significantly more affected by dietary lipids and increasing the lipid supply in feed leads to increased fat deposition and growth (Watanabe 1982). Even though there seems to be low or no positive correlation between growth and weight (Kause *et al.* 2002; Quillet *et al.* 2005; Jiang *et al.* 2015), it has been reported that excessive fat can affect fillet quality (Gjedrem 1997) and depress growth due to limited ability to digest and absorb a high amount of lipid (Du *et al.* 2005; Luo *et al.* 2005). The selection for fishes high lipid content without compromise for depressed growth becomes a challenging task in aquaculture as the levels of fat in the muscle correlates with attributes of flavour and texture (Robb *et al.* 2002). In order to understand the underlying cause of such variation in lipid content and/or composition, recent studies have been focused on transcriptomic studies to fill in the gaps in fish lipid metabolism. For example, in rainbow trout, lipid metabolism genes such as peroxisome proliferator-activated receptor (*ppar*), fatty acid binding protein (*fabp*) are have been reported to be significantly different between fishes with high or low fat content and similar results were observed between big and small fishes even under different environments (Danzmann *et al.* 2016; Hu *et al.* 2016a). Studies comparing fed and starved fish have been carried out in fishes such as European seabass (Rimoldi *et al.* 2016), gilthead seabream (Benedito-Palos *et al.* 2014), channel catfish (Preziosa *et al.* 2013) and hybrid

tilapia (Han *et al.* 2011) to understand the plasticity of lipid-related genes and its effect on other growth-related genes. Additionally, nutrigenomics studies have also been carried out to understand the effects of diets in fishes such as Atlantic salmon (Tacchi *et al.* 2011; Morais *et al.* 2012), sole (Campos *et al.* 2010), turbot (Peng *et al.* 2014), rainbow trout (Kolditz *et al.* 2007), Nile tilapia (Peng *et al.* 2016) and grass carp (Tian *et al.* 2015a) in order to elucidate how dietary changes can affect the lipid deposition, gene regulation, protein turnover in the flesh.

1.5 Tilapia biology and genetics

1.5.1 Biology

Tilapia, also known as the "aquatic chicken" is a common name for over 70 species of fishes and belongs to the large order of Perciformes, in the family of Cichlidae. Phylogenetic studies separated the tilapiines into three genera, *Tilapia* (substrate spawners), *Sarotherodon* (paternal or biparental mouthbrooders) and *Oreochromis* (maternal mouth brooders) based on their breeding behaviour and feeding habits (Klett & Meyer 2002). Tilapiines are native to Africa and the Middle East but the vast majority of its production comes from mostly Asia instead of its native range. Tilapia has become the second most important group of farmed fish in the world with a production of more than 5.3 million tonnes in 2014 (FAO 2014). Out of the three genera, the *Oreochromis* species, namely *O. niloticus*, *O. aureus*, *O. mossambicus* and *O. honorum* are the most significant species in aquaculture production (Webster & Lim 2006; McAndrew *et al.* 2016). The main cultured species globally is the Nile tilapia (*O. niloticus*) with a global production of over 3.9 million tonnes in 2015 (FAOSTAT 2017). A number characteristics make tilapia excellent candidates for culture: (i) highly adaptable in terms of feed formulations and has efficient food conversion ratios (Jauncey 2000; Ng & Romano 2013), (ii) highly adaptable to different environments (El-Sayed 2006), (iii) year-round spawning (Phillips & Kohler 1991; Beardmore *et al.* 2001), (iv) faster growth

rate and market demand (Kobayashi *et al.* 2015). Tilapia are commonly cultured in semi-intensive systems and have been genetically improved and farmed in more than 120 countries, increasing the efficiency of production as tilapias can reach market size (i.e. 600-900g) in 6-9 months of culture (Webster & Lim 2006; Yue *et al.* 2016).

1.5.2 Genetic improvements via selective breeding

The introduction of genetic improvements began in Asia in the 1990s and have been ongoing ever since (Gjedrem *et al.* 2012; Bentsen *et al.* 2017). However, these genetic programmes are mainly focused on growth and intensification hence, tilapia culture faces many problems such as disease and environment (Wang & Lu 2016). There have been several selective breeding programmes for Nile tilapia; Genetic Improvement of Farmed Tilapias (GIFT), GET-EXCEL FaST, GST (GenoMar Supreme Tilapia) and Hainan Progift (Ponzoni *et al.* 2011). Among all these, the decade-old GIFT project is the most successful and best documented (Ponzoni *et al.* 2011; Gjedrem 2012; Bentsen *et al.* 2017). It has been reported that Asian tilapia stocks were of poor genetic quality and were probably suffering from genetic bottlenecks due to poor broodstock management, thus the GIFT programme is family-based using traditional selective breeding as an approach for genetic enhancement to ensure continued improvement with maximum benefits (Acosta & Gupta 2010). To date, improvements in tilapia cultures have been mainly on growth rate, but other economical traits such as flesh composition and characteristics, salt and cold tolerance, colour, and production of all-male offspring are also a focal point in tilapia breeding.

Studies conducted have reported that genetic gains in harvest weight range from 10 to 15 percent over six generations (Ponzoni *et al.* 2011) and fillet weight had increased more than 15 percent over three spawning seasons (Nguyen *et al.* 2010b). Even though the change in flesh characteristics (i.e. protein, fat and moisture percentages, pH and colour) were negligible (Hamzah *et al.* 2016), the fatty acid composition in the flesh of quick growing strains resulted in positive changes, which may be beneficial to humans

(Nguyen *et al.* 2010c). Salinity tolerance in tilapia found that least tolerant was Nile tilapia and the most tolerant was black-chinned tilapia, being able to survive up to 120 parts per thousand (ppt) salinity (Lemarié *et al.* 2004). More than 20 inter-specific hybrids and GIFT strains were tested over ten different environments of varying salinities from 0 to 42 ppt, based on their growth and survival, and broodstock populations were produced based on the improvements of one or more traits (Tayamen *et al.* 2002). As for cold tolerance, studies have shown that different strains vary in their susceptibility to cold temperatures and is directly related to geographical regions as well as the size (Cnaani *et al.* 2000; Atwood *et al.* 2003). It has been reported that Nile tilapia from GIFT programme (Sifa *et al.* 2002) and *O. mossambicus* was the most susceptible to cold when compared to *O. aureus*. In addition, *O. aureus* was reported to be one of the strains most resistant to cold and had positive genetic correlation response to selection for higher growth rate (Cnaani *et al.* 2000; Thodesen *et al.* 2013). Studies conducted to improve the colour pigmentation of tilapia (i.e. red coloration) had also yielded significant improvements after four generations (Thodesen *et al.* 2013). Lastly, studies conducted on monosex production with a preference on all-male cultures showed that males are superior in growth rate compared to mixed-sex (Mair *et al.* 1995; Beardmore *et al.* 2001; Lind *et al.* 2015). The decreasing land space for freshwater aquaculture has highlighted the importance of salt-tolerant tilapia strains/hybrids for aquaculture. Previous studies on tilapia have all been conducted in freshwater due to the slow growth of saline tilapia when compared to the Nile tilapia (Kamal & Mair 2005), therefore it would be essential to genetically improve the former, easing the burden on the land space whilst exploiting the vast resource of the sea.

1.5.3 Genomic tools in tilapia

1.5.3.1 Molecular markers and linkage mapping

The availability of genomic tools and the advancing molecular and sequencing technologies have enhanced selective breeding of tilapia. Microsatellite markers and SNPs are currently two of the more commonly used marker systems. They have been useful in many applications including genetic analysis of populations for conservation (Bezault *et al.* 2011), pedigree assignments (Trọng *et al.* 2013), development of genetic maps (Lee *et al.* 2005), phylogenetic analysis for differentiating between species (Xia *et al.* 2014b; Syaifudin 2015) and understanding selection signatures (Xia *et al.* 2015).

In order to enhance the improvement of traits, genetic linkage maps that are required for QTL mapping and analysis are constructed. The first genetic map of *Oreochromis spp.* using haploid progeny and 174 markers was constructed (Kocher *et al.* 1998) and later refined using an F₂ interspecies hybrid population (Lee *et al.* 2005). The physical map was also constructed using cytogenetics methodology with repetitive elements (Martins *et al.* 2004) and BAC (bacterial artificial chromosome) libraries (Katagiri *et al.* 2005). Several studies were also constructed to provide higher resolution linkage maps including linkage maps from salt-tolerant hybrid species (Liu *et al.* 2013a) using microsatellites and Nile tilapia (Guyon *et al.* 2012) using radiation panels.

1.5.3.2 Genome sequencing and transcriptomics

Recently, the sequencing of the Nile tilapia genome was completed and assembled (Brawand *et al.* 2014). The Nile tilapia genome was reported to be 1.01 gigabase (Gb) instead of 1.06 Gb (Majumdar & McAndrew 1986) and distributed over 22 chromosome pairs with 24,559 protein-coding genes, more than 250 microRNA (miRNA), including seven novel miRNA and approximately 20% of the genome are transposable elements (Brawand *et al.* 2014). Additionally, a large collection expressed sequence tags (ESTs) with 17,722 annotated sequences were generated to annotate the genomic assembly (Lee

et al. 2010). Even though the improvement of the current genome assembly is still in progress (Conte *et al.* 2017), the rapid advancement and use of multiple sequencing technologies for increasing genetic data and transcriptome, will enhance the assembly and be one of the most important genomic tool that will significantly enhance the improvement of traits in tilapia (McAndrew *et al.* 2016). For example, using the genome assembly as a reference, selection signatures were identified using 1.43 million SNPs with linkage disequilibrium (LD) block size ranging from 10-100 kilobase (kb) in 47 tilapia individuals, differing in geographic locations were reported, allowing for a better understanding of the effects of selection (Xia *et al.* 2015). Mitochondrial genomes of several species of tilapia were reported to be 16 kilo basepairs (Kbp) in size and structured similarly to other cichlids, including 13 protein-coding genes, two rRNA and 22 tRNA genes and one control region have also been published (He *et al.* 2011; Kinaro *et al.* 2015). Studies have used this variability as an indication of genetic diversity in broodstock populations (Zou *et al.* 2015) and genetic purity of strains (Ordoñez *et al.* 2016). Studies conducted in Atlantic salmon have reported that mitochondrial variability can affect traits due to selection powers (Consuegra *et al.* 2015) but this has yet to be reported in tilapia. Nonetheless, with the growing publications based on NGS technologies of tilapia transcriptomes under different conditions, this void is slowly filling up. RNA-seq was utilised by several studies to understand traits including sex determination (Tao *et al.* 2013), environmental stress including salinity (Lam *et al.* 2014; Xu *et al.* 2015) and alkalinity (Zhao *et al.* 2015), disease (Zhang *et al.* 2013; Zhu *et al.* 2015; Wu *et al.* 2016) and lipid metabolism due to diet changes (He *et al.* 2015). Additionally, the complex ancestry of red tilapia (a hybrid of *O. mossambicus* and other species) was also uncovered through SNP discovery from RNA-Seq data by (Xia *et al.* 2014b). Riding on the advancing wave of NGS technology, small noncoding RNAs (ncRNA), microRNAs (miRNAs) that plays crucial roles in regulating gene expression and cell function (Bartel 2005) have been studied on in many teleost (Bizuayehu & Babiak 2014). In tilapia, there have been several reports of miRNAs regulating trait

expression including growth (Huang *et al.* 2012; Yan *et al.* 2012a; Yan *et al.* 2013a), susceptibility to environmental stress (Yan *et al.* 2012b; Yan *et al.* 2013c; Zhao *et al.* 2014; Zhao *et al.* 2016), effects of dietary additives (Tang *et al.* 2013) and skin pigmentation (Yan *et al.* 2013b). Additionally, more than 20 homologous miRNA to other species and its targets were identified *in silico* by (Li *et al.* 2015b), which has been known to regulate gene expression of traits. With the rapid advancement of NGS technologies, the otherwise information which has taken much longer time to reveal is within grasp and this massive influx of information will no doubt prove to be useful in improving breeding.

1.5.3.3 QTL mapping for economical traits

In QTL mapping, the number, location, and effect of the loci associated with the traits are important, and candidate genes or molecular markers resulting from this could be useful in MAS to improve in aquatic production. In tilapia, several traits associated with genomic loci or genes have been reported using different genetic maps, including growth, sex, salinity tolerance and lipid composition are of particular interest in this thesis and will be further discussed below. For growth, QTL have been reported in different hybrids. (Cnaani *et al.* 2004) reported that QTL detected in LG 1, 3 and 23, explaining the phenotypic variance of 5.5% to 19.3% to be associated with body weight and length in F₂ hybrid crosses of *O. mossambicus* X *O. aureus*. This supported his previous study (Cnaani *et al.* 2003) where LG 23 was associated with growth and cold tolerance. In both studies, markers in the QTL associated with growth were not homogenous due to resolution and sampling design. In a second study conducted to understand the underlying causative factor of differing growth, using a population involving two species, *O. mossambicus* and red tilapia, two major QTL was detected on LG 1 and LG 7 (Liu *et al.* 2014a). One major QTL detected on LG 1 explained 71.7% of the phenotypic variation in body weight and this linkage group has been reported to be sex-linked (Cnaani *et al.* 2008) and supported the previous study conducted by (Cnaani *et al.* 2004).

Based on the reference genome of Nile tilapia, growth hormone receptor 2 (*ghr2*) was identified as a candidate gene in the other major QTL on LG 7. It was shown that different alleles from different species (one from Nile tilapia and another from *O. mossambicus*, Mozambique tilapia) play a role in directly influencing the insulin growth factor 1 gene (*igf-1*) expression, causing growth variation among the tilapia species.

Sex determination in tilapia is rather complex as both XX/XY and ZW/WW systems are known to occur due to diverse genetic backgrounds (Campos-Ramos *et al.* 2003; Lee *et al.* 2004; Cnaani *et al.* 2008; Liu *et al.* 2013a). Nonetheless, monosex cultures, especially all-male culture, have been preferred in aquaculture production due to better growth performance (Mair *et al.* 1995; Beardmore *et al.* 2001; Toguyeni *et al.* 2002). Using genetic maps, QTL associated with sex have been associated with three linkage group, LG 1, LG 3 and LG 23/LG 22. The first reports to map sex in tilapia was conducted by (Lee *et al.* 2003) and (Ezaz *et al.* 2004) and phenotypic sex in Nile tilapia was associated with LG 8 and physically mapped onto the long arm of chromosome one. With the increase in resolution of the linkage maps generated using hybrid species, the sex-determining locus was mapped onto several linkage groups, LG 1, 3 and 22/23 (Cnaani *et al.* 2003; Cnaani *et al.* 2004; Lee *et al.* 2004; Lee *et al.* 2005; Cnaani *et al.* 2008; Liu *et al.* 2013a). It has also been reported that LG 2, 6 and 23 was associated with a specific genotype in the gender and sex-specific mortality (Shirak *et al.* 2002). Using NGS technology, the sex-determining regions of Nile tilapia on LG 1 have been narrowed down to a 2 cM interval, corresponding to 1.2 Mbp (million basepairs) on the Nile tilapia genome (Palaikostas *et al.* 2013; Brawand *et al.* 2014). Further studies have also identified a candidate gene, *Osfand3* on LG1 where differing genotypes are associated with the sex (Ma *et al.* 2016a). Two sex-determining regions have also been identified in LG 23 (Shirak *et al.* 2006), harboring anti-Mullerian hormone (*amh*) and double sex and mab related transcription factor (*dmrta*) genes that have been known to determine sex in fish species (Herpin & Schartl 2011; Pfennig *et al.* 2015). Duplication in the *amh* gene denoted as *amhy* and a missense SNP was reported to be essential for

male sex determination in Nile tilapia (Eshel *et al.* 2014; Li *et al.* 2015a). Through fine-mapping, a male-specific allele was identified in a marker (i.e. UNH898) that was significantly associated to sex and individuals lacking in this allele, resulted in all-female progeny (Eshel *et al.* 2011). Additionally, one out of five markers developed for the production of all-male progeny has yielded positive results even though there was no mention of the growth performance of the individuals (Sun *et al.* 2014b).

Besides sex determination, salinity tolerance has also been associated with the growth performance of tilapia (Watanabe *et al.* 1985; Kamal & Mair 2005). Salinity tolerance varies greatly among the species, with Nile tilapia and Mozambique tilapia being the least and most tolerant, respectively, indicative of a genetic control for salinity tolerance (Stickney 1986; Doudet 1992; Suresh & Lin 1992; Avella *et al.* 1993; Gonzalez 2012). Offspring of hybrids of species that can survive in saline waters (*O. mossambicus*, *O. aureus*) have been reported to be highly salt-tolerant and it was reported that culturing hybrids in saline conditions increases the growth rate of these hybrids (Watanabe *et al.* 1988), while increasing salinities inhibited growth in Nile tilapia (Likongwe *et al.* 1996). Increasing the knowledge of salinity tolerance without a reduction in growth performance and health will enhance production and selection for salt tolerance. There have been limited studies on the mapping of salt tolerance in tilapia. Several candidate genes (beta haemoglobin, Ca^{2+} transporting plasma membrane ATPase, pro-opiomelanocortin and beta-actin) have been mapped onto three linkage groups (LG 4, 17 and 23) through differential gene expression studies in Nile tilapia (Rengmark *et al.* 2007). Transferrin (Rengmark & Lingaas 2007) and prolactin 1 gene have previously been reported to be associated with salt tolerance were mapped onto LG 18 and LG 4 (Lee *et al.* 2005). Recent transcriptomic studies conducted using NGS technologies in both Nile (Avarre *et al.* 2014; Xu *et al.* 2015) and Mozambique tilapia (Lam *et al.* 2014) have shown us the impacts of varying salinities on gene expression in different signalling pathways, and further expanded this repertoire of candidate genes for salt tolerance. To date, there is neither any direct and positive association between

lipid metabolism/composition and salinity nor much QTL mapping for lipid composition conducted in tilapia. Even though (El-Sayed *et al.* 2005) have reported the positive effects of dietary lipids on reproductive performance in Nile tilapia reared in both freshwater and seawater, there was no significant correlation to increased growth (Nguyen *et al.* 2010c).

1.6 Aim and outline of this thesis

Tilapia is ranked as the second most cultured fish species globally. Tilapia tolerant to high salinity would greatly increase the supply for declining fish stocks by expanding the range of production in many parts of the country and regions of the world and several studies have shown the potential of tilapia in the saline environment (Kamal & Mair 2005). Current studies have been mainly focused on freshwater Nile tilapia (Uchida *et al.* 2000) and selection for its rapid growth and disease resistance (Wang *et al.* 2016). Breeding for the improvement of its lipid composition, mainly omega 3 fatty acid content and its effects from a nutrition viewpoint has not yet begun. A major concern in the consumption of tilapia is the low content of omega-3 and an elevated amount of omega 6 fatty acids, which may promote inflammation (Young 2009). To date, there has been limited information on salt tolerant tilapia and selecting it for quick growth and increased lipid composition in the flesh. Mapping QTL for improved growth rate and better lipid composition in the flesh and identifying possible candidate genes are, essential in understanding the underlying cause of this phenotypic variation to improve MAS in salt tolerant tilapia. With the advancement of sequencing technology, an investigation of the transcriptome using extreme phenotypes in a population between fast and slow growing fish will allow us to narrow down on the possible candidate genes and their effects. Additionally, a genetic map with higher resolution will be constructed to narrow down the regions on the chromosome that may harbour genes that play a role in the regulation of growth and lipid composition in the flesh of salt tolerant tilapia.

Therefore, the aim of this study is to identify the DNA markers associated with growth and fatty acid contents in salt tolerant tilapia to accelerate the genetic improvement of these traits, and to understand more about the molecular mechanism underlying the phenotypic variations in growth and fatty acid contents.

The specific objectives of the thesis are summarized as below:

- 1) To conduct preliminary QTL mapping for growth and sex in salt tolerant tilapia
- 2) To generate a large number of genome-wide SNPs using the NGS-based approach of GBS
- 3) To construct a high-density linkage map of salt tolerant tilapia
- 4) To conduct high-resolution QTL mapping for growth and omega 3 fatty acid content
- 5) To sequence mRNA from fast and slow growing salt tolerant tilapia using RNA-seq
- 6) To assemble reference transcriptome of salt tolerant tilapia
- 7) To identify DEGs between fast and slow growing salt tolerant tilapia
- 8) To identify candidate genes that lie within or are in close proximity to the identified QTL.

The following methods and procedures are utilized to achieve the abovementioned objectives:

1. Microsatellite marker genotyping for preliminary QTL mapping
2. Construction of genetic map using Joinmap 4.0 for preliminary QTL mapping
3. Conducting preliminary QTL mapping for growth and sex using MapQTL 6.0
4. Selecting 100 fast and 100 slow growing fishes to generate genome-wide SNPs for genotyping through genotyping-by-sequencing (GBS)
5. Constructing higher resolution genetic map using Joinmap 4.0 for QTL mapping using both microsatellite and SNP markers

6. Conducting QTL mapping for growth and lipid composition (i.e. total lipid and fatty acid content) in flesh of salt tolerant tilapia using MapQTL 6.0
7. Collecting RNA samples from fastest and slowest growing fishes cultured in seawater
8. Sequencing mRNA using Illumina HiSeq 2000
9. Construction of reference transcriptome of salt tolerant tilapia (inclusive of Nile tilapia mRNA using Trinity platform and CAP3 software
10. Identifying DEGs for growth and biosynthesis of unsaturated fatty acids (PUFA) by Partek and Blast2GO
11. Identifying DEGs located in identified QTL for growth and fatty acid contents

The data generated from the QTL mapping, the differentially expressed candidate genes located in the QTL will enhance our understanding of the genetic architecture of growth and lipid biosynthesis/metabolism in salt tolerant tilapia. In addition, the markers associated with these important traits could be utilized in the MAS for growth and lipid composition in the flesh of salt tolerant tilapia to accelerate genetic improvement.

1.7 Project roadmap

Figure 1.9. Flowchart of experimental design for the discovery of QTL, candidate genes and genetic markers for growth and fatty acid content in salt tolerant tilapia.

Chapter 2. Materials and Methods

2.1 Ethics statement

In this thesis, the handling of all tilapia fish was conducted in accordance with the guidelines on the care and use of animals for scientific purposes set up by the Institutional Animal Care and Use Committee (IACUC) of the Temasek Life Sciences Laboratory (TLL), Singapore. The TLL's IACUC had specifically approved this study within the project "Breeding of Tilapia" (approval number TLL (F)-12-004). All fishes in the project were anaesthetized using Aqui-S® (New Zealand) for ease of handling during trait measurement and fin clip collection. The fishes were anesthetized by submerging them in Aqui-S® at a concentration of 10 mg/L of seawater for approximately 10-30 minutes depending on size and age of fish. For all fish in QTL mapping for fatty acid and growth (QTL-omega) and RNA sequencing project, the fishes were euthanized by an overdose of Aqui-S® at a concentration of at least 100 mg/L in the for over 20 minutes until total loss of gill movement.

2.2 Mapping population, salinity acclimatization, and fish husbandry

One Mozambique tilapia male and one red tilapia female were used to generate an F₁ family. The Mozambique tilapia male weighing 521 g, was randomly selected from an F₂ generation that originated from a wild F₁ population in South Africa. The red tilapia female weighing 723 g, was a genetically improved (quick growth and color) hybrid strain of unknown genetic background and was obtained from a local farm in Singapore. One F₁ male and one F₁ female were randomly selected and crossed to generate an F₂ family. The F₂ mapping population consisted of 522 offspring and was cultured in the fish facility of Temasek Life Sciences Laboratory. At 17 days post-hatch (dph) in 200 L tanks, the fingerlings were transited from freshwater to seawater over a period of one week by adding 30 parts per thousand (ppt) of filtered seawater at a flow-rate of 3.7 L/Hour and maintaining the ammonium levels of the water below 0.5 mg/L and pH at 8.0. During the period of transition, the fingerlings were fed to

satiation three times daily with commercial feed containing 41.7% protein, 22.1% fat, 3.5% fiber, 1% calcium, 0.9% phosphorus, and 0.3% sodium (Biomar, Nersac, France). When the fingerlings were fully transitioned into 30 PPT seawater, they were transferred into 3-ton tanks using a flow-through system situated at the Marine Aquaculture Centre, Singapore and maintained on a strict feeding regime with the same feed till trait collection at 180 dph.

For the QTL mapping for fatty acid content and profiling project, the commercial feed for the fishes were changed post trait collection at 180 dph to commercial feed containing 48% protein, 16% fat, 3.0% fiber, 1.1% phosphorus, and 7.3% ash (Biomar, Nersac, France) till harvest at 220 dph. Different stages of fishes require different types of feed. In this study, salt tolerant tilapia was first bred in freshwater and subsequently acclimatized and cultured in seawater. Therefore, the feeding regime and type of feed used was an admix in order to ensure proper development during the 90-180 dph, juvenile phase and maximal growth performance and good health status from 180dph onwards, which is the grow-out phase leading to harvest. The feed that was given to the fishes in the study adhered to commercial husbandry methodology of culture hence the different feeds were given during different phases (Davis 2015).

2.3 Measuring growth traits and collection of fin clips and dorsal muscle tissue

Body weight (BW), total length (TL) and body thickness (BT) were recorded at 180 dph for the 522 F₂ individuals. BT was measured as the biggest width across the lateral line of the fish (Rutten *et al.* 2005) (Figure 2.1). The fishes were tagged individually with a passive integrated transmitter tag (8 mm PIT Tag, Green Tag Pte Ltd, Singapore) containing digital numbers to allow for the future collection of trait data and identification. Pearson's correlation coefficients between traits were calculated using statistical functions in Microsoft Excel.

Figure 2.1. Morphometric measurements obtained from F₂ mapping population.

Total length (TL) and body thickness (BT) were measured in millimetres. A portion of dorsal muscle collected was boxed in red.

Dorsal fin clips of both parents and all offspring were sampled, preserved in absolute ethanol and stored in -20 °C for DNA isolation and downstream analyses. Dorsal muscle of the 356 F₂ individuals belonging to the fastest and slowest fishes in the population was collected at 220 dph and stored in -80 °C for total lipid extraction and fatty acid profiling and quantification through fatty acid methyl ester (FAME) extraction.

2.4 Visual sexing of tilapia

All tilapia reaching 180 dph were sexed through manual observation of the genital pore (Figure 2.2). The trait data on sex was recorded with binary code; male (trait value of 1) and female (trait value of 0). Fishes with the unclear distinction of sex or underdeveloped reproductive organs were also recorded.

Figure 2.2. Genital papilla of tilapia.

Adapted from (Webster & Lim 2006)

2.5 Determining total neutral lipid content and profiling fatty acids

2.5.1 Extraction and quantification of total neutral lipid

An absolute quantification of the total neutral lipids was carried out using bligh and dyer (Bligh & Dyer 1959) from the dorsal muscles of each fish ($n = 356$) in triplicates. Briefly, 10 g of muscle sample was freeze-dried with a freeze-dryer (Virtis, Warminster, USA) for 72 hours. The samples were prepared for lipid extraction and fatty acid analysis by manually grinding them using a mortar and pestle (Figure 2.3). The ratio of dry sample weight to fresh weight was calculated for an estimation of water content in each sample. Approximately 0.25 g of ground tissue was placed into a screw-capped Pyrex culture tube and five milliliters of methanol/chloroform (2:1 vol/vol) was added. The samples were mixed gently at room temperature for 20 mins and spun at 2700 rpm for 10 mins and the supernatant was collected. Chloroform/methanol/water to a final ratio of 8:4:3 is added to the supernatant. The organic phase was transferred to a pre-weighed tube and the extracted lipids (in triplicates) were determined gravimetrically by drying them at 60 °C for 24 hours.

Figure 2.3. Preparation of samples for lipid and fatty acid extraction through freeze-drying and grinding of the dorsal muscle of tilapia.

2.5.2 FAME extraction and quantification and qualification of fatty acids

For the quantification of fatty acids, a total of 246 individuals with the highest and lowest body weight were selected. Direct fatty acid methyl esters (FAME) synthesis of fatty acids was carried out in triplicates using the ground dorsal muscles previously prepared (Figure 2.3) using a simplified protocol by O'Fallon (O'Fallon *et al.* 2007). Briefly, 0.5g of the ground sample was placed into a screw-capped Pyrex culture tube and one milliliter of the C13:0 internal standard (0.5 mg of C13:0/mL of MeOH), 0.7 mL of 10 N KOH in water, and 5.3 mL of MeOH were added into the tube. The tube was incubated in a 55 °C water bath for 1.5 hours. After cooling below room temperature, 0.58 mL of 24 N H₂SO₄ in water was added and incubated again in a 55 °C water bath for 1.5 hours. Three milliliters of hexane was added to extract the FAME products. The clear hexane layer was stored in -20 °C in GC vials before gas chromatograph (GC) analysis. The fatty acids of the FAME were determined by using the Shimadzu GC-MS QP2010 (Nakagyo-ku, Kyoto, Japan) equipped with an HP-88 capillary column, 30 m X 0.25 mm X 0.2 µm (Agilent Technologies Inc., Santa Clara, CA). The operating conditions were: injector temperature: 250 °C, detector temperature: 200 °C, helium flow rate: 1.08 mL/min. The

initial oven temperature was held at 140 °C for 3 minutes, subsequently increased to 240 °C at a rate of 7 °C/min and then held for 7 minutes. The split ratio was 10:1. For each run, Supelco™ 37 component FAME Mix standards prepared in accordance to manufacturer's instructions were placed as the first and last sample and subjected to the same run conditions to ensure the quality of the GC-MS procedure. Any disagreement of the two spectral outputs will deem the run unsatisfactory for further analyses.

Sixteen of the most common and abundant fatty acids were identified using the NIST08 library (Stein & Scott 1994) and their retention times were compared with those of the Supelco™ 37 component FAME Mix standards (Sigma, USA). Each fatty acid was reported as the percentage of the total peak area of the 16 most common and abundant fatty acids. The quantification (mg/mL) of fatty acids were calculated based on the internal standard, tridecanoic acid (C13:0, Sigma, USA), which had a known concentration of 0.5 mg/mL. The quantification reported was an average of the triplicates per sample. Pearson's correlation coefficients between traits were calculated using statistical functions in SPSS statistical software (v20).

2.6 Preparation of genomic DNA

2.6.1. Genomic DNA isolation for QTL scan for growth and sex

Genomic DNA (gDNA) of each fish was isolated with a simplified method established in the laboratory (Yue & Orban 2005). Briefly, approximately 100 mg of dorsal fin clip tissue was cut and ethanol was removed as much as possible. The tissue was homogenized in an extraction buffer containing proteinase K (final concentration of 100 µg/µL) (Sigma-Aldrich, Missouri, USA) and incubated at 55 °C and 180 rpm for two hours. Sample lysate was centrifuged at 4000 rpm for 2 min. A 96-well Pall filter column plate (Pall Corporation, Washington, USA) was used for purifying the gDNA in the supernatant. In order to activate the silica, 90 µL of 6M NaI was added to the plate and incubated for 5 minutes prior to the addition of 30 µL of the supernatant. The plate with

the mixture was incubated at room temperature for 2 min before centrifugation at 4000 rpm for 2 min. Genomic DNA bound to the column was cleaned by adding 240 μ L of washing buffer and centrifuged at 4000 rpm for 2 min. The column was left to dry to remove the remaining ethanol from the wash buffer at room temperature for 5-10 min. Reverse osmosis (RO) water (120 μ L) was added to the plate and left to stand at room temperature for 5-10 min at room temperature before centrifugation at 4000 rpm for 5 min for the elution of the gDNA. The concentration of the gDNA was determined using Nanodrop (Thermo Fisher Scientific, Waltham, USA) and the gDNA was electrophoresed on 2% agarose gel stained with 1% ethidium bromide. The gel was visualised using Gel DOC XR (Bio-Rad Laboratories, Hercules, USA) and the gDNA was diluted to a final concentration of 10 ng/ μ L as a working stock and stored in -80 °C for downstream analyses.

2.6.2. Genomic DNA isolation for QTL mapping for fatty acid

contents

The extraction of gDNA for a panel of 356 F₂ individuals were carried out using the high salt precipitation method (Aljanabi & Martinez 1997). This is because previous gDNA extraction, an approach undertaken in our lab (Yue & Orban 2005) extracted gDNA of a lower quality due to degradation and protein contamination, and did not meet the requirements for construction of sequencing libraries since the later required high-quality gDNA free of protein contamination and were minimally degraded.

Briefly, approximately 100 mg of dorsal fin clip was cut and ethanol was removed as much ethanol as possible. The fin clip was homogenized in 800 μ L of salt homogenizing buffer (0.4 M NaCl, 10 mM Tris-HCl pH 8.0 and 2 mM EDTA pH 8.0) containing 8 μ L of proteinase K (final concentration of 400 μ g/ml) (Sigma-Aldrich, Missouri, USA) and incubated at 55 °C and 180 rpm for two hours. After which, 600 μ L of 5M NaCl was added to the lysate and vortex at maximum speed for a maximum of 30

sec. The lysate was then centrifuged at 14000 g for 30 min at 4°C and supernatant transferred to a clean tube, where an equal volume of ice-cold isopropanol was added. The mixture was incubated at -20 °C for 1 hour and centrifuged at 14000 g for 20 min at 4 °C. A pellet should be observed and the supernatant decanted. To wash gDNA pellet, 1 ml of 70% ethanol was added and centrifuged at 14000 g for 10 min at 4 °C. Lastly, ethanol was removed and gDNA pellet was air-dried at room temperature prior to adding 100 µL of DNase-free water. The gDNA was stored at -20 °C for future analyses. The gDNA quality was assessed in two parts. Firstly, 1 µL of DNA was electrophoresed on 2% agarose gel stained with 1% EB and visualized using Gel DOC XR (Bio-Rad Laboratories, Hercules, 1453 USA) and its concentration determined using Nanodrop (Thermo Fisher Scientific, Waltham, USA). Secondly, the concentration of the gDNA was determined by utilizing the plate reader of Infinite® 1455 M1000 PRO (Tecan, Männedorf, Switzerland) and Qubit® dsDNA HS Assay Kit (Life 1456 Technologies, Carlsbad, USA). The protocol was carried out following the manufacturer's instructions. Briefly, a 200 times dilution of the dsDNA HS reagent was carried out using the dsDNA HS buffer and this constitutes the working solution. Three known concentration of internal standards were prepared by mixing 200, 195 and 190 µL of working solution with 0, 5 and 10 µL of standard solution, respectively. For the samples of gDNA, 198 µL of working solution was mixed with 2 µL of extracted gDNA. All these were carried out in a 96- well assay plate which was run on the plate reader of Infinite® M1000 PRO for signal detection. The signal detected was converted to a value of concentration based on the three known internal standards. Finally, gDNA was diluted to a final concentration of 20 ng/µl as a working stock and kept in -80 °C for downstream analysis.

2.7 Preparation of RNA

2.7.1 RNA isolation and assessment of quantity and quality

At 110 dph, four fishes were randomly selected each, from the fast and slow growing fishes. Fishes were dissected and samples from five different tissues, namely muscle (M), liver (L), brain (B), gills (G) and intestine (I) were collected and frozen in liquid nitrogen and stored in -80 °C for further downstream analyses. Total RNA was extracted using RNeasy® Mini kit (Qiagen, Hilden, German) according to the manufacturer's instructions. In brief, 20-30 mg of sample was homogenized in 300 µL RLT buffer using a power homogenizer before adding 300 µL more RLT buffer to a final volume of 600 µL. The lysate was centrifuged for 3 min at 14000 g at 4 °C and the supernatant was carefully removed by pipetting into a new RNase free tube. To the supernatant, 1 volume of ice-cold 75% ethanol was added and mixed well by pipetting and 700 µL of this mixture was transferred to the RNeasy Mini spin column placed in a 2 mL collection tube. The column with the mixture was centrifuged at 8500 g for 15 sec at 4 °C and the flow through was discarded. To the column, 700µL of Buffer RW1 was added and centrifuged at 8500 g for 15 sec at 4 °C. The flow-through was discarded and 500 µL of Buffer RPE was added to the column as a washing step. This was centrifuged at 8500 g for 15 sec at 4 °C and repeated for the second time but was centrifuged at 8500 g for 2 min at 4°C. The flow-through was discarded and the spin column was placed onto a new collection tube and centrifuged at 14000 g for 1 min to dry the column. The spin column was then placed onto a new 1.5 mL tube and 40 µL of RNase free water was added directly to the spin column. The column was centrifuged at 8500 g for 1 min at 4 °C to elute the RNA. In order to maximize the yield of the RNA extracted, the eluate from the tube was added onto the spin column again and centrifuged at 8500 g for 1 min at 4 °C for the second time. The eluted RNA was separated into two tubes for storage and further downstream analyses. In order to assess the RNA quality, a ten-time dilution was carried out by adding 9 µL of RNase free water to 1 µL of RNA and this working solution was

electrophoresed on 1% agarose gel stained with 1% ethidium bromide (EB) and visualized using Gel Doc XR (Bio-Rad Laboratories, Hercules, USA). The concentration of the RNA was then determined by two methods. Firstly, the concentration was assessed by adding 1 μ L of the working solution onto the arm of Nanodrop (Thermo Fisher Scientific, Waltham, USA). Secondly, the original RNA solution was assessed with Agilent 2100 Bioanalyzer (Agilent Technologies, Santa Clara, USA) following manufacturer's protocol. In brief, all the RNA samples were diluted to a final concentration ranging from 50 to 500 ng/ μ l based on previous Nanodrop results. A new RNA chip was primed on the chip priming station according to manufacturer's instruction. The priming of the RNA chip was carried out by adding 9 μ L of gel-dye mix to the right well and moving the plunger to the position of 1 mL following the closing of the chip priming station. The plunger was pressed till it was held by the clip and it was held for 30 sec before releasing the clip and waiting for 5 sec. The chip was then released by slowly moving the plunger back to the 1 mL position. 5 μ L of RNA marker and 9 μ L of the gel-dye mixture were added to another two well while 1 μ L of the prepared ladder was added to the well labelled for the ladder. Lastly, 1 μ L of RNA was added to the sample wells. The chip was vortexed at 2400 rpm for 1 min and checked and removed of bubbles prior to placing the chip into the Agilent 2100 Bioanalyzer. Only samples with RNA integrity number (RIN) above 7.5 were used for subsequent steps. The whole procedure resulted in a total of 40 total RNA samples from the fast and slow growing group: M_{Fast}, L_{Fast}, B_{Fast}, G_{Fast}, I_{Fast} and M_{Slow}, L_{Slow}, B_{Slow}, G_{Slow}, I_{Slow}. According to quantity and quality assessment from Agilent 2100 Bioanalyzer, one mg of RNA from each group were pooled before being submitted to Macrogen, Inc, South Korea for pair-end RNA-sequencing using Illumina Hiseq 2000 platform.

2.8 QTL mapping - Genotyping

2.8.1 Genotyping of microsatellites for QTL for growth and sex

From a previously published linkage map (Liu *et al.* 2013a), a total of 144 polymorphic microsatellite markers covering the 22 linkage groups (LG) of the tilapia genome near-evenly were selected and fluorescently labelled (Appendix 1 and 2). Fluorescent-labelling primers increase the throughput of the genotyping procedure and all the forward primers were labelled with FAM and HEX fluorophores. PCR amplification was carried out for each sample in a 25 µL reaction volume containing 10 ng genomic DNA, 1 x PCR buffer containing 1.5 mM MgCl₂, 0.2 µM dNTPs, 0.2 µM forward and reverse primers, and 0.5 units of Taq DNA polymerase (Finnzymes, Espoo, Finland). The PCRs were performed in PTC-100 thermal cycler (MJ Research, CA, USA) under the following conditions: 2 min denaturation at 94 °C; 35 cycles of 30 s at 94 °C, 30 s at 55–60 °C and 30 s at 72 °C and a final extension at 72 °C for 10 min. Five microliters of PCR product was electrophoresed on 2% agarose gel stained with 1% EB and their concentration was estimated by comparing the brightness of the bands with the commercial DNA ladder. Generally, ten to twenty PCR products were diluted and mixed together to increase the genotyping output and decrease time and cost. The mixed PCR products were resolved using capillary electrophoresis on an ABI3730xl DNA Genetic Analyzer (Applied Biosystems, Foster City, USA), where fragment sizes were analyzed against the internal size standard of GeneScan-500 ROX (Applied Biosystems, Foster City, USA) using GeneMapper 4.1 (Applied Biosystems, Foster City, USA). These genotypes were exported to an Excel table for further data analysis.

2.8.2 Genotype by sequencing (GBS) library preparation

The library preparation was carried out using the double digestion RAD-seq method (Peterson *et al.* 2012) with some modifications. Each gDNA (200 ng) was digested with

20 units of restriction enzyme of *PstI-HF* and *MspI* (New England Biolabs, Ipswich, MA, USA) at 37 °C for 2.5 hours and were ligated with barcoded adaptors (Peterson *et al.* 2012), pooled, size selected using Pippin Prep (Sage Science, Beverly, MA, USA) for fragments of 350 to 600 bp, and purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). These size-selected libraries were PCR amplified using Phusion® High-Fidelity DNA Polymerase (New England Biolabs, Ipswich, MA, USA), and the PCR- enriched libraries were further purified using QIAquick PCR Purification Kit (Qiagen, Hilden, Germany). Prior to sequencing, the libraries were quantified using KAPA Library Quantification Kits (Kapa Biosystems, Wilmington, MA, USA) by qPCR in the My iQ Thermal Cycler (Bio-Rad Laboratories, Hercules, CA, USA). The libraries were sequenced on the NextSeq 500 platform (Illumina, San Diego, CA, USA) to generate raw sequencing single-end reads of 150 bp in length.

2.8.3 Processing of NGS reads, identifying and genotyping SNPs for scan of QTL for growth and fatty acids

Low quality reads and reads with any uncalled base were removed by the program *process_radtags* implemented in Stacks package (version 1.21) (Catchen *et al.* 2011). All the clean reads were trimmed to 125 bp to rule out sequencing error at the ends of reads, followed by a final step of de-multiplexing bioinformatically and assigning clean reads to each sample. In brief, clean reads were mapped to the reference genome of Nile tilapia OreNil2.0 using the program BWA (Li & Durbin 2009) with a maximum mismatch of 5. The alignments with unique mapping targets were further used for stacks assembly using the program *pstacks* (-m, 3) under the SNP model and the assembled stacks from parental samples were used to construct a catalogue of loci using the program *cstacks* in Stacks package, with default parameters. SNP discovery was carried out by mapping the assembled stacks back to the catalogue loci using the program *sstacks*. The program *genotypes* was used for SNP genotyping across this mapping

family (Catchen *et al.* 2011). A minimum of five times sequence coverage was applied for each sample and any loci with more than 20% missing data were removed from further analysis.

2.9 Construction of linkage maps

2.9.1 Linkage map construction using microsatellites

For the construction of linkage maps, the F₂ population consisting of 522 individuals of 180 dph was separated to four groups, all offspring, male offspring, female offspring, as well as male and female offspring. All genotyping data from 144 microsatellite markers were formatted in accordance with JoinMap 4.0 (Van Ooijen 2006) using the standard codes for “cross-pollination” (CP) type progeny. The genotypes were checked for inconsistencies with Mendelian inheritance using Chi-square test performed using JoinMap 4.0 and manually corrected for error. Samples with more than four missing genotypes and deviate from the expected ratio of 1:1 were removed. The linkage amongst the markers was evaluated with the Kosambi mapping function for the ‘best-fit’ map distance in the analysis and markers with the logarithm of odds (LOD) > 3.0 for segregation data were indicative of a linkage and assigned to a linkage group (Van Ooijen 2011). All linkage groups were numbered in correspondence to the published map (Liu *et al.* 2013a) based on selected markers. All linkage maps were visualized using MapChart 2.1 software (Voorrips 2002). The total length of the linkage map, average marker distance and the female to male recombination ratio were calculated using statistical functions in Microsoft Excel. In order to conduct QTL mapping for sex, a linkage map was constructed using the 399 F₂ individuals with known sex. The trait data was recorded with binary code; male (trait value of 1) and female (trait value of 0). The construction and visualization of the linkage map were conducted with the same parameters as mentioned above. The all-offspring linkage map was a sex-averaged map and was constructed without reference to the sex of the individuals while the male and

female offspring linkage maps, similarly to the all-male and all-female linkage maps were constructed only with individuals with known sex.

2.9.2 Linkage map construction using microsatellites and SNP

For the construction of linkage maps, the F₂ population consisting of 356 individuals of 180 dph of known sex was used. All the SNP and microsatellite markers were submitted to JoinMap 4.0 (Van Ooijen 2006) to construct a genetic map following the procedure described above. However, the linkage amongst the markers was evaluated with the Kosambi mapping function for the 'best-fit' map distance in the analysis and markers with the logarithm of odds (LOD) ≥ 7.0 for segregation data were indicative of a linkage and assigned to a linkage group.

2.10 QTL mapping

2.10.1 QTL mapping for growth traits and sex

Post construction of the four linkage maps using microsatellite markers, QTL mapping for growth traits was performed using MapQTL 5.0 software (Van Ooijen 2004) for all offspring, male offspring, and female offspring, respectively. Simple interval mapping (SIM) analysis (Lander & Botstein 1989) based on maximum likelihood (ML) approach was first used to identify the segregating QTL at each position of the genome and the genome was scanned at 1 centimorgan (cM) intervals between loci. This is done to identify markers flanking the putative QTL that are significantly associated with the trait to constitute a set of cofactors. Using these markers as cofactors will reduce genetic background variation that was induced by QTL located elsewhere in the genome (Jansen & Stam 1994) and increase the sensitivity and accuracy of QTL detection. The selection of cofactors to be used in the analysis was carried out using the automated cofactor selection analysis where a backward elimination procedure was applied to the initial set of cofactors and only significant markers at $p < 0.02$ were used as cofactors. Multiple QTL

mapping (MQM) (Jansen 1993) was then carried out using these cofactors and scanning the genome at 1cM intervals. In order to determine the statistical significance of the QTL identified, permutation tests were carried out. Permutation tests were carried out based on the empirical distribution values of randomizing trait values with respect to the genotypes was used to determine the significant Log of Odds (LOD) thresholds under the null hypothesis of no segregating QTL (Churchill & Doerge 1994). The permutation test was performed with 10,000 re-iterations to determine the LOD threshold across single chromosomes as well as genome-wide. QTL detected were deemed as suggestive if the LOD scores fell between 5% to 1% at the chromosome-wide level, and significant if the LOD scores fell below 1% at chromosome-wide or fell below 5% at genome-wide level (Lander & Kruglyak 1995). QTL mapping for sex was carried out using the SIM based on ML analysis to using MapQTL 5.0 as previously described.

2.10.1.1 Fine mapping of significant QTL for growth in male offspring

Primer sequences of microsatellites flanking the detected significant QTL for growth traits on LG 2 and 18 were extracted and NCBI-blastn was performed against the assembled tilapia genome (Brawand *et al.* 2014) to identify the genome sequences flanking these QTL. From the genomic sequences extracted, 18 additional microsatellites were identified and primers for these additional 18 microsatellite markers (Appendix 2) were designed using Primer 3 (Untergasser *et al.* 2007). All 235 male offspring in the F₂ family were genotyped and the LG 2 and LG 18 of the male genetic map was reconstructed to include the additional markers. Repeated QTL analysis was performed on a genome-wide scale as described above.

2.10.2 QTL mapping for total lipid and fatty acid content (Omega 3 and 6)

Post construction of linkage map using both microsatellite and SNP markers, MAPQTL 6.0 software (Van Ooijen 2004) was used to detect QTL associated with the phenotypic variations of total lipid and fatty acid content in the dorsal muscle of the fish. QTL mapping for total lipid and fatty acid content was carried out using the SIM based on ML analysis to using MapQTL 5.0 as previously described.

2.11 Transcriptomic analysis of fast and slow growing fish

2.11.1 RNA-Sequencing library preparation

Ten sequencing libraries from the five tissues of the fast and slow growing groups were constructed. One µg of total RNA for each pooled sample was used for library construction using TrueSeq™ RNA sample prep kit (Illumina, San Diego, USA). In brief, the total RNA was treated with DNase to remove the genomic DNA contamination, and mRNA with poly (A) was enriched using magnetic beads coated with oligo-dT subsequently. The enriched mRNA was fragmented and cDNA was reverse-transcribed with random hexamer-primers to produce second strand cDNA. The second strand cDNA was purified and ligated with Illumina sequencing index and adapters. Size selection for the cDNA was carried out and enriched to generate the final sequencing libraries. Sequencing of these libraries was then performed on the Illumina HiSeq™ 2000 (Illumina, San Diego, USA) to produce paired-end (2 X 101 bp) sequence reads.

2.11.2 Processing of sequencing reads and *de novo* assembly of reference transcriptome

The Illumina short raw reads were processed by NGS QC toolkit (Patel & Jain 2012) to remove adapters and low-quality reads ($Q < 20$) and unpaired reads. The filtered high-quality reads were *de novo* assembled using Trinity (version 20140717) (Grabherr *et al.*

2011; Haas *et al.* 2013). As Nile tilapia is closely related to the hybrid tilapia in this study, the *de novo* assembly was merged with 45,440 annotated Nile tilapia mRNA sequences from NCBI RefSeq database using CAP3 software (Huang & Madan 1999) to maximize the chance of presenting as many transcripts as possible. The merged transcriptome was BLASTN-searched against the Nile tilapia mRNA to assign gene descriptions, by using an E-value cut-off of 1×10^{-6} and selecting the best hit. The reference transcriptome used for subsequent analyses comprised of the set of unique BLASTN-annotated sequences, where a single sequence was selected as a representative for each corresponding Nile tilapia BLASTN match, as well as the remaining unannotated sequences. A BLASTX-search against the NCBI RefSeq protein database was also further carried out for the unannotated sequences to supplement the annotated dataset. The assembly of the reference transcriptome is outlined in Figure 2.4.

Figure 2.4. Flowchart of transcriptome assembly and annotation of salt tolerant tilapia.

2.11.3 Identification of differentially expressed transcripts in the muscle and liver

As this study is focused on the growth-related and lipid biosynthesis related differences in the muscle and liver, the filtered reads from the muscle and liver samples (M_{Fast} , L_{Fast} , M_{Slow} , and L_{Slow}) were aligned to the reference transcriptome using the CLC Genomics Workbench (version 7.0), 'Map reads to reference' tool where the thresholds are set at 95% length fraction, 95% similarity fraction with default costs for mismatches, insertions, and deletions. The BAM files of the mapping were then imported into Partek Genomics Suite (version 6.6) for differential gene expression analysis. Following mRNA quantification to obtain read counts and reads per exon kilobase per million mapped reads (RPKM), differentially expression analysis was performed using ANOVA. As there were no replicates in this study, the p -value was adapted from the mRNA quantification step where the algorithm estimates a p -value by assuming that all the samples are replicates and that the transcripts are evenly distributed across all samples. The p -value was subsequently corrected using the Bonferroni method (Haynes 2013) available in the software. The differential expression analysis was first narrowed down to transcripts with BLASTN or BLASTX annotations. Secondly, the global RPKM values were examined to filter off very lowly expressed transcripts, and $RPKM > 1$ was set as the minimum threshold for expression (Figure 2.5). Annotated transcripts were considered differentially expressed if the corrected p -value < 0.001 and fold change ≥ 2 .

2.11.4 Functional classification and pathway analysis of differentially expressed transcripts (DETs)

Gene Ontology (GO) is an international classification system that comprehensively profiles characteristics of genes, gene products and sequences across species. It allows us to have an overview of the different functional classes in our differentially expressed

transcripts that may underlie superior growth and possible differing lipid content between the fast and slow growing fishes. To understand the possible role of the differentially expressed transcripts (DETs) in conferring the size and lipid content, the annotation of gene ontology (GO) terms was carried out using Blast2GO (Götz *et al.* 2008) based on the GO annotations of its corresponding homologs in the NCBI RefSeq database allowing for an overview of the functional classes. These DETs were then mapped onto KEGG pathways using KAAS webserver (Moriya *et al.* 2007) to allow us to have an overview of the network of pathways that were involved in the differing size and lipid content between the fast and slow growing fishes. From the assigned KEGG pathways, 18 and 21 pathways of interest were selected for growth and lipid biosynthesis.

Figure 2.5. Frequency distribution of reads per exon kilobase per million mapped reads (RPKM) values of differentially expressed transcripts in dorsal muscle and liver of salt tolerant tilapia.

2.11.5 Reverse transcription to convert RNA to single-stranded cDNA (sscDNA)

Prior to reverse transcription, gDNA contamination in the RNA samples was removed using DNase I (Roche, Basel, Switzerland) following manufacturer's instructions. In

brief, a total reaction volume of 25 μ L, containing 2.5 μ L of 10X incubation buffer, 2 units of DNase I, 2 units of RNase inhibitor (Promega, Madison, USA), 4 μ g of total RNA and an appropriate amount of RNase-free water was prepared. This reaction mix was first incubated at 37 °C for 20 min to digest the gDNA and then incubated at 75 °C for 20 min to inactivate the DNase. Reverse transcription was then carried out using iScript™ reverse transcription supermix (Bio-Rad Laboratories, Hercules, CA). Briefly, a reverse transcription master mix of 180 μ L was prepared by adding 48 μ L of iScript RT super mix to 132 μ L of nuclease-free water. The reaction volume is 20 μ L, containing 15 μ L of the master mix, 1 μ g of RNA and an appropriate amount of nuclease-free water. The reaction is then incubated at 25 °C for 5 mins to prime the reverse transcriptase, then incubated at 46 °C for 20 min to start the reverse transcription and lastly, incubated at 95 °C for 1 min to inactivate the reverse transcriptase. The sscDNA were diluted five times and kept in -20 °C for further analyses.

2.11.6 Real-time quantitative reverse transcriptase-PCR (qRT-PCR) for validation of expression of differentially expressed transcripts (DETs)

A total of 21 primer pairs (17 up-regulated and 4 down-regulated) were designed using Primer- blast software (Ye *et al.* 2012) for both muscle and liver transcripts that were differentially expressed between the fast and slow growing fishes (Appendix 3). Quantitative PCR (qPCR) was carried out on Biorad iQ5 (Bio-rad laboratories, Hercules, USA) using SYBR Green as a fluorescent dye. qPCR reactions were carried out in triplicates and a 20 μ L reaction volume containing 2 μ L (10 ng) of 10 times diluted sscDNA, 0.2 μ L of each primer of concentration 10 μ M, 5 μ L of 2x master mix from KAPA SYBR® FAST qPCR Kits (Life Technologies, Carlsbad, USA) and 2.6 μ L of sterile water. Raw data was converted to cycle threshold (Ct) values using the software provided by Biorad iQ5 (Bio-Rad Laboratories, Hercules, USA). Quantification of

relative gene expression was calculated using the $2^{-\Delta\Delta Ct}$ method (Livak & Schmittgen 2001) using beta-actin as the housekeeping gene to normalize the relative expression. Paired t-test was used for calculation of p -values for comparing the differential expression of genes between fast and slow growing fishes. The fold change of each gene was the ratio of fast growing fish to that of slow growing fish and all results were compared with RNA sequencing fold change data.

2.12. Differentially expressed genes with sequence variation that are located within or with close proximity of the QTL regions for growth and fatty acid content

In order to understand the potential impact of DEGs, co-localization of QTL with the DEGs from the RNA sequencing data will facilitate the identification of candidate genes underlying the phenotypic variation in the population.

2.12.1 Genomic location of DEGs with identified QTL regions for growth and fatty acid content through linkage and physical mapping

From the QTL identified for growth and fatty acid content, the focus was placed on the three growth-related traits, BW, TL and BT and omega 3 fatty acid content. In order to physically locate the genomic regions of the flanking markers where these QTL lie, the microsatellite or SNP marker sequences were downloaded from NCBI database and the NGS data from the SNP marker discovery. The genomic locations of the markers and DEGs were identified using the BLASTN program against the Nile tilapia genome (Brawand *et al.* 2014). DEGs that lie or are within close proximity to the identified QTL regions were selected for further study.

2.12.2 Tissue expression of DEGs

Total RNA from seven tissues, muscle, liver, gills brain, intestine, kidney, and heart was extracted and reverse transcribed to cDNA as described in section 2.7.2 and 2.7.3. Nine primer pairs were designed using Primer- blast software for qPCR to allow us to understand the relative gene expression of each DEG in different tissues (Appendix 13). The qPCR was carried out in triplicates and following the procedure as described in section 2.11.6.

2.13 Statistical analyses

The common statistical analyses used in the correlation of traits, QTL mapping analyses, and differential gene expression are listed below.

The correlation of traits was conducted and its significance was investigated by using Pearson's correlation two-tailed test (Section 2.3 and 2.5). This test was performed with IBM SPSS statistical software package (v20).

Chi-square goodness of fit test ($p < 0.05$) was calculated for the segregation of each DNA marker using the JoinMap software for the construction of the linkage maps (Section 2.9). The LOD thresholds and significance of the QTL mapping for the traits (i.e. growth and fatty acid content) were investigated by conducting permutation tests with the MapQTL software (Section 2.10).

Lastly, the Bonferroni methodology of correcting p values was applied to identify differentially expressed genes (DEGs) between fast and slow growing fishes. Genes were considered differentially expressed if the corrected p -value < 0.001 and fold change ≥ 2 . One-way ANOVA with Tukey's HSD test using the IBM SPSS statistics software package (v20) was conducted to verify the DEGs from the RNA-seq as well as on the differential expression of selected DEGs in different tissues. Results differential expression of genes in different tissues were expressed as mean $\pm$ SD and for the verification of the DEGs, they were compared with RNA-seq data (Section 2.11 and 2.12).

Chapter 3. Results

QTL mapping and the identification of candidate genes for growth and fatty acid content for salt tolerant tilapia have been conducted. Firstly, trait measurements have been conducted and analysed and DNA markers have been discovered and genotyped. Secondly, the construction of linkage maps, both microsatellite-based and microsatellite and SNP based have been carried out and QTL mapping has been conducted for growth and fatty acid content in F₂ salt tolerant tilapia. Lastly, differential gene expression has been identified between fast and slow growing individuals through the assembly of a reference transcriptome and these differentially expressed genes were found to be located within or within close proximity of the identified QTL. Details of the results of these analyses will be described below.

3.1 Trait values and their correlations

3.1.1 Growth traits

In order to perform QTL mapping, the trait values for three growth-related traits, body weight (BW), total length (TL) and body thickness (BT) were collected from an F₂ full sib mapping population consisting of 522 individuals. The sex of the fishes was recorded with high stringency where unclear or underdeveloped genitaliae were recorded as well. In the population, 235 and 164 individuals were male and female respectively, while for the remaining 123 individuals, the sex could not be determined due to underdeveloped genitaliae. Analyses of the phenotypic data clearly showed substantial levels of phenotypic variation in all three traits (Table 3.1). The average body weight, total length and body thickness of all offspring were 270.2 ± 70.8 g, 223.4 ± 20.6 mm and 61.3 ± 7.2 mm, respectively. The average body weight, total length and body thickness of the male offspring were 291.3 ± 66.0 g, 231.6 ± 17.7 mm and 63.5 ± 6.9 mm, respectively. The average body weight, total length and body thickness of the female offspring were 239.2 ± 63.8 g, 213.0 ± 19.2 mm and 59.7 ± 7.0 mm. Additionally, all traits displayed

qualities of a quantitative trait as their distribution resembles a bell-shaped curve (Figure 3.1). The correlation between body weight and total length ($r=0.93$, $p<0.01$) was strong, while the moderate correlation between body weight and body thickness and total length and body thickness was observed ($r=0.7$, $p=0$ and 0.69 , $p=0$) (Table 3.2).

Table 3.1. Averages growth traits for all, male and female offspring recorded in 180 dph (days post-hatch) in an F₂ full-sib mapping family of salt-tolerant tilapia

	Combined	Male	Female
Trait values	(n = 522)	(n = 235)	(n = 164)
Body weight (g)	270.2±70.8	291.3±66.0	239.2±63.8
Total Length (mm)	223.4±20.6	231.6±17.7	213.0±19.2
Body Thickness (mm)	61.3±7.2	63.5±6.9	59.7±7.0

Figure 3.1. Frequency distribution of three growth traits in salt tolerant tilapia. Body weight, total length, and body thickness were measured in 522 F₂ offspring.

Table 3.2. Correlations between recorded growth traits of salt tolerant tilapia at 180 dph

	180BW	180TL	180BT
180BW	-	0.93	0.70
180TL		-	0.69
180BT			-

Pearson's correlation was conducted and all correlations were statistically significant. ($p<0.05$)

3.1.2 Total neutral lipid and fatty acid content

The measurement of the contents of total neutral lipid and fatty acids was carried out using gravimetry and GC-MS respectively, hence the absolute quantification and percentages of the total neutral lipid and 16 fatty acid components in wet weight could be determined and is presented in Table 3.3. The mean wet weight of neutral lipids and fatty acids isolated were 17.3 ± 6.1 mg/g and 10.2 ± 4.9 mg/g respectively. Of the 17 fatty acids identified in the GC-MS, oleic acid (C18:1), palmitic (16:0) and docosahexaenoic acids (C22:6) are the major fatty acids, accounting for $20.7\% \pm 3.5$, $18.3\% \pm 4.6$ and $16.4\% \pm 3.0$ of total fatty acids, respectively. PUFA represented the biggest fraction of total fatty acid composition at 42.0%, comprising of 25.0% and 17.1% of omega 3 (PUFA n3) and omega 6 (PUFA n6) fatty acids, followed by SFA at 30.0% and MUFA at 27.6% (Table 3.3).

3.1.2.1 Correlations between fatty acids

In general, there was positive correlation amongst the fatty acids (Table 3.4) that were all statistically significant ($p < 0.01$). Moderate to strong correlation coefficients were observed between saturated fatty acids (SFA) (C14:0, C16:0 and C18:0) and monounsaturated fatty acids (MUFA) (C16:1, C18:1, C20:1), ranging from $r=0.74$ to $r=0.86$. In contrast, C14:0 was observed to be strongly correlated ($r > 0.8$) to C18:2n6, C18:3n3, C20:3n6 and C22:5n6 while weakly to moderately correlate to the rest of the fatty acids. Similarly, C16:0 and C18:0 was strongly correlated ($r > 0.8$) to C18:2n6 and C22:6n3 and C18:2n6, C18:3n3, C20:4n6, C20:5n3, C22:5n6 and C22:5n6 respectively, while being weakly to moderately correlate to the rest of the polyunsaturated fatty acids (PUFA). The MUFA were generally moderately to strongly correlated with the PUFA ranging from $r=0.62$ to $r=0.96$ with the exception of C18:1 and C18:3n6 where there was a weak correlation ($r=0.35$). The PUFA were on the whole moderately to strongly correlated to each other, ranging from $r=0.68$ to $r=0.9$. However, it has been observed

that C18: 3n6 is moderate to weakly correlated ($r < 0.5$) with all PUFA as well as all other fatty acids with the exception of C16:1 and C18:2n6.

3.1.2.2. Correlation between fatty acids and growth traits

The correlation between the fatty acids and body weight or total length was statistically significant ($p < 0.05$) and presented in Table 3.5. In general, all fatty acids were not significantly correlated with body thickness. The phenotypic correlations between the SFA (C14:0, C16:0 and C18:0) and body weight (BW) or total length (TL) and between MUFA (C16:1, C18:1, C20:1) and BW or TL exhibited weak positive correlations with r values ranging from 0.177 to 0.231 and 0.091 to 0.237, respectively, even though C20:1 was not significantly correlated with TL. All PUFA were mostly weakly positively correlated to BW or TL with r values ranging from 0.119 to 0.202 and 0.110 to 0.184. Both Omega 3 fatty acids (PUFA n3) and omega 6 fatty acids (PUFA n6) were weakly positively correlated with two growth traits (BW or TL) with r values ranging from 0.150 to 0.182.

Table 3.3. Phenotypic values of total neutral lipid (mg/g), fatty acid content (mg/g wet weight), fat and moisture content in the flesh of 246 salt tolerant tilapia individuals

Trait	Common name	Type	Mean mg/g	%	Standard deviation mg/g	%	Maximum mg/g	%	Minimum mg/g	%
<i>Fatty acid composition</i>										
C14:0	Myristic acid		0.3	2.7	0.2	0.9	0.9	9.0	0.0	0.6
C16:0	Palmitic acid		1.9	18.3	1.1	4.6	5.8	23.8	0.2	4.9
C18:0	Stearic acid		0.9	9.2	0.4	1.7	2.8	13.4	0.1	1.9
Total SFA			3.1	30.3	1.5	4.2	8.6	38.3	0.3	18.7
C16:1	Palmitoleic acid	n-7	0.6	5.4	0.4	1.6	2.0	15.9	0.0	1.1
C18:1	Oleic acid	n-9	2.2	20.7	1.3	3.5	7.6	30.2	0.2	5.1
C20:1	Eicosenoic acid	n-9	0.2	1.6	0.1	0.8	0.6	11.6	0.0	0.6
Total MUFA			2.9	27.6	1.7	4.4	10.1	37.7	0.2	11.2
C18:2	Linoleic acid (LA)	n-6	1.0	9.3	0.5	1.4	3	20.1	0.1	5.2
C18:3	γ -Linolenic acid (GLA)	n-6	0.1	0.6	0	0.6	0.5	6	0	0.1
C18:3	α -Linolenic acid (ALA)	n-3	0.1	1.1	0.1	0.3	0.4	2.7	0	0.3
C20:2	Eicosadienoic Acid	n-6	0.1	0.7	0	0.3	0.3	2.6	0	0.2
C20:3	Dihomo-gamma-linolenic Acid (DGLA)	n-6	0.1	0.8	0	0.3	0.4	3.4	0	0.3
C20:4	Arachnoic acid (AA)	n-6	0.4	4.3	0.2	1.2	0.9	8	0	1.6
C20:5	Eicosapentaenoic acid (EPA)	n-3	0.2	2.4	0.1	0.7	0.5	5.1	0	0.8
C22:4	Adrenic acid	n-6	0.1	1.3	0.1	0.5	0.4	3.4	0	0.6
C22:5	Docosapentaenoic acid (DPA)	n-3	0.5	5.1	0.2	1.1	1.6	13.8	0	2.7
C22:6	Docosahexaenoic acid (DHA)	n-3	1.6	16.4	0.7	3.0	3.3	30.6	0	0.9
Total PUFA			4.2	42.1	1.8	5.0	9.8	55.4	0.4	31.5
PUFA (n-3)		n-3	2.5	25.0	1.0	3.5	5.1	32.4	0.3	10.3
PUFA (n-6)		n-6	1.7	17.1	0.8	2.3	4.8	29.2	0.2	12.7
<i>Fat and moisture content</i>										
Total neutral lipid (TL) (mg/g)			17.3		6.1		34.5		5.5	
Total fatty acid (TFA) (mg/g)			10.2		4.9		28.5		1	
n6/n3 ratio*			0.7		0.6		9.6		0.5	
Moisture content (%)			76.8		1.9		80.6		66.2	

Table 3.4. Correlations between fatty acid compositions of salt tolerant tilapia (356 individuals) at 220 dph

	C14:0	C16:0	C16:1	C18:0	C18:1	C18:2n6	C18:3n6	C18:3n3	C20:1	C20:2n6	C20:3n6	C20:4n6	C20:5n3	C22:4n6	C22:5n6	C22:6n3
C14:0		0.81	0.86	0.81	0.86	0.83	0.40	0.84	0.81	0.58	0.80	0.68	0.75	0.77	0.90	0.75
C16:0			0.74	0.75	0.87	0.80	0.32	0.74	0.74	0.59	0.64	0.71	0.72	0.68	0.78	0.81
C16:1				0.79	0.85	0.96	0.71	0.78	0.78	0.62	0.65	0.64	0.76	0.73	0.76	0.74
C18:0					0.85	0.84	0.35	0.82	0.77	0.63	0.73	0.84	0.80	0.73	0.84	0.90
C18:1						0.90	0.35	0.86	0.87	0.63	0.68	0.80	0.75	0.75	0.85	0.89
C18:2n6							0.68	0.83	0.84	0.72	0.69	0.74	0.81	0.78	0.80	0.82
C18:3n6								0.34	0.38	0.46	0.38	0.24	0.46	0.46	0.33	0.28
C18:3n3									0.82	0.83	0.73	0.85	0.89	0.88	0.89	0.81
C20:1										0.66	0.67	0.74	0.74	0.73	0.81	0.76
C20:2n6											0.62	0.75	0.80	0.86	0.70	0.64
C20:3n6												0.73	0.89	0.86	0.90	0.85
C20:4n6													0.89	0.85	0.86	0.91
C20:5n3														0.86	0.86	0.84
C22:4n6															0.90	0.75
C22:5n6																0.85
C22:6n3																

Pearson's correlation was conducted and all correlations are statistically significant ($p<0.05$)

Table 3.5. Correlations between fatty acid and growth traits of salt tolerant tilapia

Fatty acid composition	BW[#]	TL[#]	BT[#]
C14:0	0.231**	0.189**	0.029
C16:0	0.219**	0.183**	-0.021
C16:1	0.237**	0.185**	0.030
C18:0	0.177**	0.156**	-0.014
C18:1	0.193**	0.149**	-0.010
C18:2n6 (LA)	0.202**	0.164**	0.003
C18:3n6 (GLA)	0.217**	0.184**	0.066
C18:3n3 (ALA)	0.159**	0.128*	-0.007
C20:1	0.114*	0.091	-0.005
C20:2n6	0.119*	0.110*	-0.008
C20:3n6 (DGLA)	0.160**	0.133*	0.016
C20:4n6 (AA)	0.126*	0.117*	-0.026
C20:5n3 (EPA)	0.140**	0.129*	-0.009
C22:4n6	0.137**	0.112*	-0.002
C22:5n6 (DPA)	0.169**	0.141**	0.004
C22:6n3 (DHA)	0.172**	0.154**	-0.042
PUFA (n3)	0.171**	0.150**	-0.028
PUFA (n6)	0.182**	0.152**	-0.001

Pearson's correlation was conducted and correlations are statistically significant if

* $p < 0.05$, ** $p < 0.01$. #BW: body weight, TL: total length, BT: body thickness

3.2 Linkage maps

3.2.1 Construction of linkage maps for all offspring, male offspring, female offspring as well as male and female offspring

Linkage maps for all offspring, only male, only female, as well as male and female offspring were constructed. The four linkage maps (Appendix 4, 5, 6, 7) concurred with the linkage map published earlier (Liu *et al.* 2013a) with 22 linkage groups (LG). From a total of 144 markers used for mapping, 125, 124, 123 and 125 markers were anchored in the all offspring, only male, only female, and male and female offspring maps, respectively. The initial construction of the only male map used 124 markers initially,

and later an additional 18 microsatellite markers (Appendix 3) were mapped on LG 2 and 18 (Appendix 5) to fine-map QTL. The total map length, number of markers and average marker distance were 1170.6 cM, 125 and 11.4 cM; 1180.6cM, 142 and 9.8 cM; 1162.1 cM, 123 and 11.5 cM as well as 1162.1 cM, 125 and 15.6 cM for all offspring, only male, only female, and male and female offspring maps, respectively (Table 3.6).

3.2.2 Comparison between male and female linkage maps

The maps of male and female offspring contained 142 and 123 markers and spanned 1180.8 cM and 1162.1 cM, respectively (Table 3.7). 116 microsatellite markers were shared between these two maps, and the total lengths of the linkage map based on shared markers were 1049.2 cM and 1051.3 cM in male and female offspring, respectively. Even though the ratio of lengths of the common interval in female and male offspring was 1.00, recombination ratios were significantly skewed towards either the female or male offspring on three linkage groups; 1.87, 0.67 and 0.67 on LG 10, LG 16, LG 17 (Table 3.7).

Table 3.6. Summary of genetic linkage map of all offspring, male offspring, female offspring and male and female offspring in salt tolerant tilapia

LG	Number of markers				Length(cM)				Average marker distance (cM) of individual linkage groups			
	All	Male	Female	Male and female	All	Male	Female	Male and female	All	Male	Female	Male and female
1	6	6	6	6	53.2	52.2	54.7	53.4	10.6	10.4	10.9	10.7
2	6	12	6	6	54.0	55.2	53.2	54.6	10.8	5.0	10.6	10.9
3	6	6	6	6	72.6	78.1	68.9	74.0	14.5	15.6	13.8	14.8
4	6	6	6	6	56.3	59.0	53.1	56.5	11.3	11.8	10.6	11.3
5	5	5	5	5	53.6	53.2	54.9	54.1	13.4	13.3	13.7	13.5
6	6	6	6	6	57.0	58.2	58.6	58.5	11.4	11.6	11.7	11.7
7	6	5	6	6	66.2	46.1	69.3	69.7	13.2	11.5	13.9	13.9
8	7	5	7	7	57.5	42.3	58.6	58.6	9.6	10.6	9.8	9.8
9	7	7	7	7	73.7	76.5	69.2	73.7	12.3	12.8	11.5	12.3
10	5	6	5	5	47.8	45.6	43.4	44.7	12.0	9.1	10.9	11.2
11	6	6	4	6	53.7	54.1	40.3	56.6	10.7	10.8	13.4	11.3
12	7	7	7	7	51.3	55.4	51.7	54.6	8.5	9.2	8.6	9.1
13	4	5	4	4	25.4	47.5	29.3	25.4	8.5	11.9	9.8	8.5
14	5	5	5	5	52.1	54.1	59.5	56.8	13.0	13.5	14.9	14.2
15	5	5	5	5	43.3	43.7	45.2	45.2	10.8	10.9	11.3	11.3
16	6	6	7	6	62.7	64.0	59.1	61.6	12.5	12.8	9.9	12.3
17	3	3	3	3	12.0	13.5	8.3	11.7	6.0	6.7	4.1	5.8
18	6	18	6	6	58.2	62.0	58.9	63.2	11.6	3.6	11.8	12.6
19	5	5	5	5	62.6	64.8	65.3	63.8	15.6	16.2	16.3	15.9
20	7	7	7	7	56.7	56.4	56.1	56.6	9.5	9.4	9.3	9.4
21	5	5	4	5	45.6	47.2	44.8	46.0	11.4	11.8	14.9	11.5
22	6	6	6	6	54.9	51.5	59.7	54.3	11.0	10.3	11.9	10.9
Total	125	142	123	125	1170.6	1180.8	1162.1	1193.5	11.4	9.8	11.5	15.6

Table 3.7. Recombination ratios between male and female genetic maps in salt tolerant tilapia

LG	Number of shared markers	Length (cM)		
		Male	Female	F:M
1	6	52.2	54.7	1.0
2	6	43.6	42.6	1.0
3	6	78.1	68.9	0.9
4	6	59.0	53.1	0.9
5	5	53.2	54.9	1.0
6	6	58.2	58.6	1.0
7	5	46.1	49.3	1.1
8	5	46.1	42.5	0.9
9	7	76.5	69.2	0.9
10	5	23.2	43.4	1.9
11	4	41.2	40.3	1.0
12	7	55.4	51.7	0.9
13	4	22.4	29.3	1.3
14	5	54.1	59.5	1.1
15	5	43.7	45.2	1.0
16	6	64.0	42.9	0.7
17	3	13.5	8.3	0.6
18	4	28.1	34.0	1.2
19	5	64.8	65.3	1.0
20	7	56.4	56.1	1.0
21	3	17.8	21.8	1.2
22	6	51.5	59.7	1.2
Total	116	1049.2	1051.3	1.0

3.2.3 Construction of a denser linkage map using GBS technology

3.2.3.1 Identification and genotyping of genome-wide SNPs

Prior to the construction of the linkage map, the processing of the genotyping of the SNP markers must be carried out. A total of 321.47 million clean reads were produced by the NextSeq500 platform after sequential quality filtering and sequence trimming, within which 6.62 and 7.75 million reads were produced for parental samples, respectively and the average clean reads for F₂ progeny was 782, 683. Using the data sets of the parental samples, a catalogue containing 377,919 loci was constructed and this was used as a reference for SNP discovery and genotyping in the mapping family. A total of

37,141 SNP markers were identified across all the samples using the program *genotypes*, of which, 927 SNPs were genotyped in more than 80% of progeny samples. All these genotyped SNPs were used for linkage mapping.

3.2.3.2. Construction of denser linkage map using microsatellite and SNP markers

In order to construct a denser linkage map, all the 927 SNPs discovered from GBS and 162 microsatellites from the previous reconstructed linkage map for all offspring containing 356 individuals were assessed for Mendelian segregation by JoinMap 4.0 (Van Ooijen 2006) before map construction. After removal of distorted markers, 633 SNPs and 151 microsatellites were successfully used to construct a genetic map. Among these markers, 72 SNPs and 6 microsatellites and 84 SNPs and 2 microsatellites were heterozygous in paternal and maternal samples, respectively, while 479 SNPs and 143 microsatellites markers were heterozygous in both parents. The sex average genetic map contained 784 markers and spanned 2076.7 cM with an average marker interval of 2.73 cM (Figure 3.2, Table 3.8). The smallest and largest gap was located at LG 3 with a length of 1.35 cM and largest was at LG 5 with a length of 4.93 cM between the markers.

Table 3.8. Summary of statistics for the linkage map constructed using microsatellite and SNP markers for QTL mapping in salt tolerant tilapia; total length, number of markers and average marker distance

LG	Mapped markers	Genetic length (cM)	Average marker interval (cM)
1	32	63.47	2.05
2	39	98.39	2.59
3	66	87.97	1.35
4	38	76.87	2.08
5	24	113.56	4.94
6	42	128.17	3.13
7	48	106.57	2.27
8	29	103.26	3.69
9	27	95.49	3.67
10	28	108.67	4.02
11	35	87.82	2.58
12	37	91.93	2.55
13	23	95.1	4.32
14	32	108.39	3.5
15	26	86.62	3.46
16	31	94.51	3.15
17	33	63.34	1.98
18	41	81.9	2.05
19	35	94.96	2.93
20	39	78.98	2.08
21	36	106.43	3.04
22	43	104.3	2.48
Total	784	2076.68	2.73

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring (continued).

There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring (continued)

There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring (continued)

There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring (continued)

There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

16

17

18

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring (continued)

There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

Figure 3.2. A linkage map of salt-tolerant tilapia constructed using 356 F₂ offspring. There are 22 linkage groups named with LG1-22. The number on the left is the distance in centiMorgans (cM). The labelling on the right are the names of the DNA markers. Previously mapped DNA markers are labelled with OMO or GM and superscripted hash '#' at the end while new microsatellite DNA and SNP markers are labeled with 2RM or 18RM and numerical, respectively.

3.3 Mapping QTL for economical traits

3.3.1 Mapping QTL using the microsatellite-based linkage map

3.3.1.1 QTL for sex

Two significant QTL for sex were detected on LG 1 and LG 22 respectively. The significant QTL detected were supported with LOD values of 13.38 and 9.89 explained 19.8% and 10.8% of the phenotypic variance respectively on LG 1 and LG 22 (Figure 3.3, Table 3.9).

Figure 3.3. QTL for sex detected on LG 1 and LG 22 in salt-tolerant tilapia.

The dotted line represents the genome-wide threshold $p < 0.05$ for QTL is 4.0. The genetic distance on each linkage group is in centiMorgans (cM)

3.3.1.2 Growth in all offspring, only male, and only female offspring

In this study, QTL analysis for growth traits was conducted in all 522 F₂ offspring, 235 males, and 164 females, respectively. In all 522 F₂ offspring, three suggestive QTL were detected on LG 20 for BW and TL, explaining 2.6 to 3.1% phenotypic variance respectively, on LG 22 for TL, explaining 3.3% phenotypic variance and LG 12 for BT, explaining 2.4 % of the phenotypic variance (Appendix 8, Table 3.9). As the effects of QTL for growth traits is somewhat small in all offspring in this study, separate analyses for QTL for growth in male and female offspring were conducted.

Table 3.9. Significant and suggestive quantitative trait loci (QTL) for growth traits and sex in all, male and female offspring of salt-tolerant tilapia

Traits	LOD threshold	QTL ID	LOD	Linkage		Peak position (cM)	PVE(%)
				group (LG)	QTL interval		
BW	4.0	BW	2.79*	20	8.8-15.6	11.8	3.1
m-BW	4.0	mBW-a	2.82*	2	0.0-15.0	15.0	4.7
		mBW-b	2.51*	15	0.0-21.3	4.0	6.3
		mBW-c	2.55*	17	0.0-2.2	0.0	4.1
		<u>mBW-d</u>	3.69**	18	51.4-56.4	55.4	6.2
		mBW-e	2.62*	22	11.0-27.0	17.0	5.8
f-BW	4.1	fBW-a	2.72*	1	32.6-45.9	35.6	11.6
		fBW-b	2.78*	6	0.0-8.6	4.0	7.8
		<u>fBW-c</u>	4.80***	8	3.3-16.1	4.3	12.1
TL	4.0	TL-a	2.21*	20	8.8-15.6	11.8	2.6
		TL-b	2.37*	22	12.6-31.5	22.6	3.3
m-TL	4.3	<u>mTL-a</u>	5.12****	2	0.0-15.0	15.0	8.4
		mTL-b	2.22*	17	0.0-2.2	0.0	3.5
		<u>mTL-c</u>	4.34***	18	51.4-56.4	55.4	7
		mTL-d	2.2*	22	11.0-27.0	15.0	4.1
f-TL	4.0	fTL-a	2.53*	1	32.6-45.9	35.6	8.1
		fTL-b	2.92*	6	0.0-8.6	3.0	7.9
		<u>fTL-c</u>	5.83****	8	3.3-16.1	4.3	14.3
BT	4.0	BT	2.34*	12	0.0-11.2	1.0	2.4
m-BT	4.1	mBT-a	3.44*	2	25.5-38.3	26.5	6.9
		mBT-b	2.81*	5	17.2-35.0	27.2	7
		mBT-c	2.77*	8	37.6-41.9	41.9	4.8
		mBT-d	3.32*	18	51.4-56.4	55.4	6.1
f-BT	4.1	fBT-a	2.61*	6	42.5-58.6	51.5	8
		<u>fBT-b</u>	3.71**	8	16.1-44.4	35.1	17.2
		fBT-c	2.63*	12	40.9-49.6	42.9	7.6
SD	4.0	<u>SD-a</u>	13.38*****	1	30.7-44.8	35.7	19.8
		<u>SD-b</u>	9.89*****	22	13.8-30.1	13.8	10.8

BW: body weight, TL: total length, BT: body thickness, SD: Sex determination. mBW: male body weight, mTL: male total length, mBT: male body thickness. fBW: female body weight, fTL: female total length, fBT: female body thickness. PVE: Phenotypic variance explained. LOD threshold scores are based on $p < 0.05$. Significant QTL are underlined and bolded. Significance: Chromosome-wide * $p < 0.05$, ** $p < 0.01$; Genome-wide *** $p < 0.05$, **** $p < 0.01$, ***** $p < 0.005$, *****($p < 0.001$).

Two significant and six suggestive QTL were detected in LG 2 (Figure 3.4), 18 (Figure 3.5) and 2, 5, 8, 15, 17 and 22, respectively in the male offspring. The significant

and suggestive QTL explained 3.5% to 8.4% of the phenotypic variance (Table 3.9). One of the two significant QTL detected on LG 18 was associated with all traits (Figure 3.5) while the other significant QTL detected on LG 2 was significantly and suggestively associated with TL and BW respectively (Figure 3.4). Another QTL detected on LG 2 was located in a different region and was suggestively associated with BT (Table 3.9). Suggestive QTL on LG 17 and LG 22 was associated with BW and TL, respectively and one and two suggestive QTL associated with BW and BT were detected on LG 15 and LG 5 and 8, respectively (Appendix 9 and Table 3.9).

Two significant and four suggestive QTL were detected in linkage groups LG 8 (Figure 3.6) and 1, 6, and 12 respectively (Appendix 10 and Table 3.9) in the female offspring. The significant and suggestive QTL explained 7.6 to 17.2% of phenotypic variance. One of the two significant QTL detected on LG 8 was associated with BW and TL, while the other significant QTL detected on a different region on LG 8 was associated with BT. Two suggestive QTL detected on LG 6 was associated with BW and TL and BT at two different locations while LG 1 and 12 was suggestively associated with BW and TL and BT, respectively (Appendix 10 and Table 3.9).

Figure 3.4. QTL for growth traits detected on LG 2 in male offspring of salt-tolerant tilapia.

The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers. The top of the graph is the LOD value of the QTL peaks and the dotted line of LOD value of 2.6 and 4.15 represents the LOD threshold for chromosome-wide and genome-wide significance of p -value < 0.05 , respectively. The boxes are representative of the traits associated with the detected QTL regions; BW - body weight, TL - total length and BT - body thickness.

Figure 3.5. QTL for growth traits detected on LG 18 in male offspring of salt-tolerant tilapia.

The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers. The top of the graph is the LOD value of the QTL peaks and the dotted line of LOD value 2.8 and 4.13 represents the LOD threshold for chromosome and genome-wide significance of p -value < 0.05 , respectively. The boxes are representative of the traits associated with the detected QTL regions; BW - body weight, TL - total length and BT - body thickness.

Figure 3.6. QTL for growth traits detected on LG 8 in female offspring of salt-tolerant tilapia.

The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers. The top of the graph is the LOD value of the QTL peaks and the dotted line of LOD value 2.56 and 4.03 represents the LOD threshold for chromosome-wide and genome-wide significance of p -value < 0.05 , respectively. The boxes are representative of the traits associated with the detected QTL regions; BW - body weight, TL - total length and BT - body thickness.

3.3.2 Improved QTL mapping using the microsatellite and SNP based linkage map

3.3.2.1 QTL for growth traits

Improved QTL analysis for growth traits was conducted in all 356 F₂ offspring using the linkage map with a higher density to have a better understanding of the phenotypic variances of growth in the population. Two significant QTL for body weight (BW) were detected on LG 2 and LG 6, respectively. The significant QTL detected were supported with LOD values of 4.7 and 4.7 and explained 5.1% and 5.2% of the phenotypic variance respectively on LG 2 and LG 6 (Table 3.10). The same QTL on LG 6 was detected to be associated with total length (TL), significantly on a chromosome level and was supported by a LOD value of 4.2, explaining 4.7 % of the phenotypic variation in the population. Two suggestive QTL were detected on LG 17 and LG 6 for TL and body thickness (BT) respectively. These suggestive QTL explained 3.7% and 3.6% of the phenotypic variance.

Table 3.10. Significant and suggestive quantitative trait loci for growth traits of salt tolerant tilapia

Traits	LOD threshold ($p<0.05$)		LOD	Linkage group (LG)	QTL interval	Peak position (cM)	PVE(%)	Significance
	GW	CW						
BW	4.6	3	4.7	2	10.1-24.7	16.1	5.1	SG
		3.2	4.7	6	40.5-46.6	46.5	5.2	SG
TL	4.6	3.1	4.2	6	40.5-46.6	46.5	4.7	SC*
		3.1	3.4	17	30.4-26.4	27.4	3.7	CS
BT	4.7	3.1	3.2	6	37.0-32.4	36.4	3.6	CS

BW: body weight, TL: total length, BT: body thickness. PVE: Phenotypic variance explained. LOD threshold scores are based on $p<0.05$. GW: LOD threshold at a genome-wide level ($p<0.05$), CW: LOD threshold at chromosome-wide level ($p<0.05$). Significant QTL are in bold. Significance: CS - Suggestive at chromosome-wide ($p<0.05$), SC - Significant at chromosome-wide ($*p<0.01$), SG - Significant at genome-wide ($p<0.05$).

Table 3.11. Significant and suggestive quantitative trait loci for total lipids, total fatty acids and fatty acid content in salt tolerant tilapia

Traits	LOD threshold ($p < 0.05$) GW	CW	Linkage group (LG)	LOD	Interval	Peak position (cM)	PVE (%)	Significance
TDW	4.6	3.7	3	4.1	30.3-31.6	31.1	6.6	CS
	4.6	3.2	4	3.6	35.0-42.4	36.8	5.8	CS
	4.6	3.1	6	3.6	37.0-46.6	40.5	5.8	CS
	4.6	3.1	8	4.8	65.1-69.6	67.1	8.3	SG
	4.6	3.0	13	4.5	0-25.4	10.8	7.8	SC**
	4.6	3.1	15	4.3	43.0-48.1	46.0	7.5	SC**
TWW	4.7	3.7	3	4.1	30.3-31.6	31.1	6.7	CS
	4.7	3.0	4	3.4	35.0-42.4	36.8	5.7	CS
	4.7	3.0	6	3.4	37.0-46.6	40.5	5.6	CS
	4.7	3.1	8	3.9	65.1-69.6	67.1	7.0	SC*
	4.7	3.0	12	3.0	59.6-63.9	61.7	5.1	CS
	4.7	2.9	13	3.5	0-25.4	10.8	6.2	SC*
	4.7	3.1	15	4.1	43.0-48.1	46.0	7.2	SC**
	4.9	3.2	16	3.4	90.3-94.5	94.2	5.2	CS
	4.9	3.3	19	3.4	85.7-93.0	91.7	5.7	CS
C14:0	4.7	3.0	8	3.9	65.1-69.6	67.1	6.0	SC**
	4.7	2.9	16	3.9	90.3-94.5	94.2	5.9	SC*
	4.7	3.4	22	4.1	43.7-47.5	44.7	6.2	CS
C16:0	4.8	3.6	3	4.2	28.0-30.3	30.0	6.5	CS
	4.8	3.0	16	3.3	90.3-94.5	93.3	5.1	CS
	4.8	3.2	21	3.4	47.2-51.7	51.2	5.2	CS
C16:1	5.0	3.0	8	3.5	65.1-69.6	66.1	5.6	CS
	5.0	3.0	16	3.1	90.3-94.5	94.2	5.1	CS
	5.0	3.2	20	3.2	70.8-72.1	71.8	5.1	CS
C18:0	5.7	2.9	11	3.5	43.5-49.8	45.5	6.1	CS
	5.7	3.5	20	5.5	16.0-17.3	16.9	9.5	SC*
C18:1	5.0	3.1	16	4.0	90.3-94.5	94.5	6.3	SC*
	5.0	3.2	19	3.8	85.6-92.9	91.7	6.6	CS
C18:2n6	4.9	3.3	19	3.5	85.6-92.9	92.7	6.0	CS
C18:3n6	11.8	4.6	20	9.4	0.0-2.0	1.0	17.0	SC*
	11.8	4.0	22	4.9	47.5-50.3	50.3	7.9	CS
C18:3n3 (ALA)	4.8	3.0	18	3.7	70.3-71.2	72.1	6.1	SC*
	4.8	3.2	20	3.4	7.8-15.2	15.2	6.0	CS
C20:1	5.6	3.1	16	4.2	0.0-7.5	0.0	7.0	SC**
	5.6	3.0	18	4.0	21.3-24.6	22.3	6.6	SC*
C20:2n6	7.8	3.0	18	3.2	70.3-71.2	72.1	5.8	CS
C20:3n6	9.4	3.3	4	5.4	73.8-76.9	76.8	8.8	SC**
	9.4	3.3	16	3.5	90.3-94.5	94.5	5.3	CS
	9.4	3.9	20	3.9	0.0-2.0	0.0	6.4	CS
C20:4n6	4.8	3.2	20	3.3	7.8-15.2	15.2	6.3	CS
C20:5n3 (EPA)	4.8	2.9	11	3.1	57.3-60.3	57.8	5.2	CS
	4.8	3.0	18	3.9	70.3-71.2	72.1	6.5	SC*
	4.8	3.3	20	4.1	7.8-15.2	15.2	7.5	SC*
C22:4n6	5.1	3.3	20	3.5	7.8-15.2	15.2	6.4	CS
C22:6n3 (DHA)	4.8	3.0	11	3.1	43.5-49.8	46.5	5.0	CS
	4.8	3.3	19	3.3	85.6-92.9	91.7	5.8	CS

TDW - Total Lipid (dry weight) TWW - Total lipid (wet weight). TFA - Total fatty acid. Omega 3 fatty acids: ALA - alpha linoleic acid, EPA - Eicosapentaenoic acid,

DHA - Docosahexaenoic acid. PVE: Phenotypic variance explained. LOD threshold scores are based on $p < 0.05$ (GW - Genome-wide, CW - Chromosome-wide). Significant at genome-wide QTL are in bold. Significance: CS - Suggestive at chromosome-wide ($p < 0.05$), SC - Significant at chromosome-wide ($*p < 0.01$, $**p < 0.005$, $***p < 0.001$), SG - Significant at genome-wide ($p < 0.05$).

3.3.2.2 QTL for total lipid and fatty acid content

QTL analysis for 18 traits related to fatty acids, including total lipid, total fatty acid and different types of fatty acids, were conducted in all 356 F₂ offspring of salt tolerant tilapia (Table 3.11). One significant QTL was detected on LG 8 for total lipid (dry weight) (TDW) which was supported by a LOD value of 4.8 and explaining 8.3% phenotypic variance in the population. This same QTL on LG 8 was also associated significantly on a chromosome level with total lipid (wet weight) (TWW) and C14:0 saturated fatty acid (SFA), was supported by LOD value of 3.9 and 3.9, respectively and explained 7.0% and 6.0% of the phenotypic variation. Additionally, 14 QTL were significantly associated (on a chromosomal level) with different fatty acid components on six LG with LOD values ranging from 3.5 to 9.4 and explaining 5.9% to 17% of the phenotypic variation. 21 suggestive QTL detected on 12 LG and explained 5% to 7.9% of phenotypic variance. In this study, particular attention has been placed in the QTL for omega 3 fatty acids, namely, C18:3n3 (ALA), C20:5n3 (EPA) and C22:6n3 (DHA) due to their importance in the diets of humans and animals. Two significant and four suggestive QTL have been detected at a chromosome level for the three fatty acids and they have been observed to overlap or be in close proximity to each other in four LG. One of the two QTL detected on LG 18 (Figure 3.7, Table 3.11) was significantly associated with ALA and EPA on a chromosomal level while the other QTL on LG 20 (Figure 3.8, Table 3.11) was significantly associated to EPA and suggestively associated to ALA on a chromosomal level. Two of the three QTL detected on LG 11 (Figure 3.9, Table 3.11) was suggestively associated with EPA and DHA on two different loci at a chromosomal level. One of the

three QTL detected on LG 19 was suggestively associated with DHA on a chromosomal level (Figure 3.10, Table 3.11).

Figure 3.7. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 18 in salt tolerant tilapia.

The dotted line of LOD value 3.0 represents the LOD threshold for chromosome-wide significance. The map position is in centimorgans (cM) and the marker positions are labelled on the top of the chart.

Figure 3.8. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 20 in salt tolerant tilapia.

The dotted line of LOD value 3.25 represents the LOD threshold for chromosome-wide significance. The map position is in centimorgans (cM) and the marker positions are labelled on the top of the chart.

Figure 3.9. QTL for omega 3 fatty acids, ALA, EPA, and DHA detected on LG 11 in salt tolerant tilapia.

The dotted line of LOD value 3.05 represents the LOD threshold for chromosome-wide significance. The map position is in centimorgans (cM) and the marker positions are labelled on the top of the chart.

- 106 -

3.4 Transcriptome analysis of fast and slow growing fish

3.4.1 Raw sequencing reads and processing

Ten cDNA libraries constructed from five tissues; namely, muscle (M), liver (L), brain (B), gills (G) and intestine (I) pooled from four individuals of fast-growing and slow-growing fish respectively were sequenced using the Illumina HiSeq 2000 sequencing platform. The sequencing produced a total of 492,458,402 paired-end (PE) reads of length 100 bp. Quality-trimming and filtering of low-quality reads resulted in 486,651,674 high-quality PE trimmed reads (98.81%) which were used for *de novo* assembly (Table 3.12).

Table 3.12. Statistics from sequencing reads of ten libraries generated for saline tilapia using Illumina HiSeq2000 sequencing

PE Library name	Type of tissue*	No. of raw reads	No. of reads after trim	Q20(%)
M _{Fast}	Dorsal muscle	42,807,120	42,316,956	98.85
L _{fast}	Liver	57,992,018	57,266,262	98.75
B _{Fast}	Brain	38,832,976	38,348,976	98.75
G _{Fast}	Gills	49,061,666	48,387,980	98.63
I _{Fast}	Intestine	43,791,188	43,192,548	98.63
M _{Slow}	Dorsal muscle	42,419,682	41,938,406	98.87
L _{Slow}	Liver	69,891,260	69,205,904	99.02
B _{Slow}	Brain	44,479,234	43,955,378	98.82
G _{Slow}	Gills	62,067,852	61,362,052	98.86
I _{Slow}	Intestine	41,115,406	40,677,212	98.93

Sequencing was conducted using total RNA isolated from five tissues of an individual fish. Four samples of the same tissue from fast and slow growing fishes are pooled together to generate a library. *Tissues include dorsal muscle (M), Liver (L), Brain (B), Gills (G), Intestine (I). Subscript Fast – fast growing fish and Slow – slow growing fish.

3.4.2 *De novo* assembly of reference transcriptome

In order to obtain a reference transcriptome of salt tolerant tilapia, the filtered reads were assembled using Trinity platform, generating 331,608 contigs. These contigs were then merged with 45,440 mRNA sequences of the Nile tilapia, a closely related species, using

CAP3 to generate consensus sequences. The total of 328,078 contigs with average contig length of 963 bp and the longest contig sized at 94,446 bp were obtained from the alignment using the CAP3 software. Gene descriptions were assigned to each contig using BLASTN against the NCBI Nile tilapia RefSeq mRNA database and this resulted in 36,680 annotated unique sequences (average length ~3.4kb). Even though 225,602 sequences (average length ~429 bp) remained unannotated, they were both merged to form the reference transcriptome for subsequent analyses, allowing for a more comprehensive representation of the transcriptome (Table 3.13, Figure 3.11). A subsequent BLASTX-search of the unannotated sequences against the NCBI RefSeq protein database provided descriptions for an additional 6,019 sequences, bringing the total number of annotated unique sequences to 42,699.

Table 3.13. Overview of reference transcriptome assembled using different software for RNA-seq sequences generated from ten libraries

Name of software	No. of contigs	Average length (bp)	Max length of contigs (bp)
Trinity	331,603	765	24,459
Cap3	328,078	963	94,446
Blastn (annotated)	36,680	767	94,446
Unannotated	219,583	428	7,627
Reference transcriptome	262,282	846	94,446

Cap3 merged assembly with RefSeq Nile tilapia sequences and Blastn data that is utilized for re-mapping of the *de novo* assembly of saline tilapia reference transcriptome. Annotation of the reference assembly is performed using Blastn and Blastx.

Figure 3.11. Length distribution of contigs from the *de novo* assembled reference transcriptome of salt tolerant tilapia.

3.4.3 Differential expression of transcripts between the muscle and liver of fast and slow-growing fish

In this study, my focus is on the growth and fatty acid content. Therefore, differential expression of the transcripts in muscle and liver of salt tolerant tilapia were presented here and discussed later. The trimmed reads from the fast and slow-growing muscle and liver libraries were mapped onto the reference transcriptome, allowing for reads with at least 90% similarity and 90% aligned length to be mapped (Table 3.14). In this study, only the BLASTN- and BLASTX-annotated transcripts were considered for the differential expression analysis. Following an RPKM cutoff of >1 , and thresholds for the differential expression set as Bonferroni-corrected p -value < 0.001 and fold change ≥ 2 , a total of 2,236 and 3,020 transcripts were found to be differentially expressed between the fast and slow-growing muscle and liver, respectively. It was observed that upon the comparison of fast growing fish against slow-growing fish, 427 and

1809 transcripts were up and down-regulated in the muscle respectively, whereas the corresponding numbers for the liver were 1808 and 1212 transcripts (Figure 3.12).

Table 3.14. Mapping efficiency and coverage of pair-end libraries of the muscle and liver of the fast and slow growing fishes

Pair-end Library	No. of Mapped reads	Mapped reads (%)	Pair-end reads (%)	Average PE read length (bp)	Coverage
M _{Fast}	40,234,353	95.08	79.97	218	25.7x
M _{Slow}	39,883,872	95.1	81.27	226	28.4x
L _{Fast}	51,695,157	90.27	83.88	154	43.9x
L _{Slow}	61,803,981	89.3	81.56	215	37.6x

Tissues include dorsal muscle (M) and Liver (L). Subscript Fast – fast-growing fish, Slow – slow-growing fish.

Figure 3.12. Differentially expressed transcripts (DET) in the muscle and liver of salt tolerant tilapia at 110 dph when comparing between fast to slow-growing fishes.

3.4.4 Functional classification of DETs

A total of 957 (42.7%) and 1726 (57.2%) differentially expressed transcripts were assigned 2088 and 3374 GO terms in the muscle and liver respectively, covering all three domains; biological process (BP), molecular function (MF) and cellular component (CC). In most of the cases, one transcript was assigned to many terms and these sequences were further characterized into primary subcategories (Figure 3.13).

In the muscle, the GO terms with the most assigned genes to the domain of CC were cell parts (GO:0044464) with 60 transcripts, intracellular (GO:0005622) with 40 transcripts and intracellular parts (GO:0044424) with 34 transcripts. For BP, 40 transcripts were assigned to a primary metabolic process (GO:00044238), 39 to cellular metabolic process (GO:0044237), and 34 to macromolecule metabolic process (GO:0043170) and lastly, for MF, 18 transcripts were involved in nucleotide binding (GO:0000166), 16 in ion binding (GO:0043167) and 15 in hydrolase activity (GO:0016787) (Figure 3.16).

In the liver, the most assigned GO terms in the CC domain were cell parts (GO:0044464) with 56 transcripts, intracellular (GO:0005622) with 36 transcripts and membrane (GO:0016020) with 33 transcripts. For BP, 40 transcripts were assigned to a primary metabolic process (GO:00044238), 39 to cellular metabolic process (GO:0044237), and 29 to macromolecule metabolic process (GO:0043170) and lastly, for MF, 19 transcripts were involved in nucleotide binding (GO:0000166), 16 in hydrolase activity (GO:0016787) and 15 in ion binding (GO:0043167) (Figure 3.13).

Figure 3.13. Top 10 functional classification of differentially expressed transcripts in the muscle (Top) and liver (Bottom).

The pathway-based analysis provides a broad view understanding of how growth and lipid biosynthesis is being regulated to confer a superior growth rate and possible higher lipid content at the molecular level. Therefore, the differentially expressed transcripts (p -value < 0.001 and fold change ≥ 2) were further annotated and classified based on KEGG categories and a total of 844 and 1234 DETs were mapped onto KEGG in the muscle and liver respectively. The annotation to KEGG via KAAS revealed diverse, but related top five subclass pathways, “metabolic pathways”, “endocytosis”, “ubiquitin-

mediated proteolysis”, “PI3K-Akt signaling pathway” and “protein processing in the endoplasmic reticulum” in the muscle and, “metabolic pathways”, “biosynthesis of secondary metabolites”, “biosynthesis of antibiotics”, “microbial metabolism in diverse environments” and “endocytosis” in the liver (Figure 3.14).

Figure 3.14.Top 20 KEGG pathways of differentially expressed transcripts in the muscle and liver of salt tolerant tilapia.

3.4.5 Validation of differentially expressed genes (DEG) using quantitative PCR (qPCR) transcripts

For the validation of the expression profile of differentially expressed genes (DEG) in the RNA-sequencing data, I have selected 10 and 11 genes expressed in the muscle and liver, respectively (Appendix 3) and conducted qPCR in triplicates. The sequence of the genes was extracted from the genome and primer pairs were design to span across 2 exons. Initial PCR using these primers gave single band products indicative of the reliability of these primers (data not shown) to be used for qPCR. The comparison of the fold changes (ratio of RPKM of fast-growing fish/ RPKM of slow-growing fish) of

the RNA-Seq and qPCR data showed a correlation for all the 21 genes tested, indicating that the expression profiling of the DEGs determined by the RNA-seq is reliable and accurate (Table 3.15).

Table 3.15. Fold changes of selected genes determined by qPCR for validation of RNA-seq

Tissue	Contig name	Gene symbol	Description of gene	Corrected Bonferroni <i>p</i> -value	Fold change RNA-seq	qPCR*	Status (up/down-regulated)	KO numbers
Muscle	Contig913	<i>smpx</i>	Small muscle protein, X-linked	0	4.74	4.23	up	NIL
	gi 542238929 ref XM_003452251.2	<i>myoz2</i>	Myozenin-2-like	0	2.42	1.84	up	NIL
	gi 542232991 ref XM_003449987.2	<i>tmod1</i>	Tropomodulin-1-like	0	2.92	2.56	up	NIL
	Contig18394	<i>Nrap</i>	Nebulin-related anchoring protein	0	2.45	2.68	up	NIL
	gi 525344062 ref NM_001279720.1	<i>myod1</i>	Myogenic differentiation 1	0	6.85	7.08	up	NIL
	Contig1994	<i>tgfb3</i>	Transforming growth factor beta-3-like	0	3.73	4.25	up	K13377
	gi 542232620 ref XM_005454618.1	<i>foxo3</i>	Forkhead box O3	0	0.29	0.29	down	K09408
	Contig19663	<i>cacnb1</i>	Voltage-dependent L-type calcium channel subunit beta-1 subunit-like	0	4.65	22.12	up	K04862
	gi 544583517 ref NM_001282894.1	<i>casp3</i>	Caspase-3	0	0.09	0.09	down	K02187
	Contig957	<i>Gadd45g</i>	Growth arrest and DNA damage-inducible protein, gamma-like	0	0.15	0.06	up	K04402
Liver	Contig15609	<i>lipin2</i>	Lipin 2	0	3.46	1.73	up	K15728
	gi 542234485 ref XM_003450527.2	<i>fabp10a</i>	Fatty acid-binding protein 10-A, liver basic-like	0	140.21	50.64	up	NIL
	Contig20500	<i>Elovl6</i>	Elongation of very long chain fatty acids protein 6-like	0	10.85	2.30	up	K10203
	Contig12096	<i>AdipoR2</i>	Adiponectin receptor protein 2-like	0	2.21	1.59	up	K07297
	Contig17010	<i>pcbd1</i>	Pterin-4-alpha-carbinolamine dehydratase-like	0	10.31	3.16	up	NIL
	gi 542192523 ref XM_003441797.2	<i>scd</i>	Stearoyl-CoA desaturase-like	0	52.03	9.46	up	K00507
	Contig14256	<i>rxrgb</i>	Retinoic acid receptor, gamma-B-like	0	3.43	3.00	up	K08526
	gi 542236588 ref XM_003451414.2	<i>Faxdc2</i>	Fatty acid hydroxylase domain-containing protein 2-like	0	14.71	10.83	up	K07750
	Contig7098	<i>abhd5</i>	Abhydrolase domain containing 5-like	0	12.25	3.41	up	NIL
	Contig7165	<i>lss</i>	Lanosterol synthase (2,3-oxidosqualene-lanosterol cyclase)	0	0.04	0.01	down	K01852
	Contig13166	<i>hmgcs</i>	Hydroxymethylglutaryl-CoA synthase, cytoplasmic-like	0	0.01	0.34	down	K01641

*One-way ANOVA was conducted for the comparison of fold change values between fast and slow growing samples and all genes are significantly differentially expressed. ($p < 0.05$).

3.4.6 Selected KEGG pathways associated with muscle development and lipid biosynthesis and the DEGs in muscle and liver

Considering the aim of the study, I have selected 18 and 21 KEGG pathways in muscle and liver, respectively, to provide an overview of the DEG and the role they may play in conferring the differing phenotypes in fast and slow growing fish muscle and liver. A total of 171 genes, of which 36 and 135 genes were up and down-regulated respectively, were differentially expressed in 18 muscle development associated KEGG pathways (Appendix 11). For the liver, a total of 308 genes, of which 169 and 139 genes up and down-regulated respectively, were differentially expressed in 21 lipid biosynthesis associated pathways (Appendix 12).

3.5 Differentially expressed genes (DEG) with sequence variation that are located inside or within close proximity of the QTL

3.5.1 Genomic locations of DEG

Sequence variation in the DNA sequences may alter the efficiency of translated proteins, more so in the 5' and 3' untranslated regions, where regulatory and signalling elements may be located (Mignone *et al.* 2002; Jackson *et al.* 2010; Barrett *et al.* 2012; Matoulkova *et al.* 2012). In this study, I have focused on the DEGs that lie or are within close proximity of the identified QTL to facilitate the identification of candidate genes and their polymorphisms that underlie the phenotypic variation of growth and fatty acid content especially omega 3 fatty acids in the population. I have identified ten growth or lipid biosynthesis related DEGs that lie or are within reasonably close proximity of the identified QTL for growth (77kb-1.8Mb) and omega 3 fatty acid content (<400kb) (Table 3.16). For growth-related QTL, four genes were found to be located within or within close proximity of the QTL region on LG 2, 6 and 7. One of the two genes, *forkhead box K1 (foxK1)* showed down-regulation in fast-growing fishes and was found

to be located approximately 1Mb downstream of both BW and TL on LG 6 while *ubiquitin specific peptidase 38 (usp38)* was located inside of the detected QTL for BT on the same LG, LG 6 was down-regulated in fast-growing fish. Two other genes, *secreted protein acidic and cysteine-rich (sparc)* and *smad family member 3 (smad3)* which were up and down-regulated in fast-growing fish muscle respectively, were located approximately 77kb and 300kb downstream of QTL identified in BW and TL respectively. (Table 3.16). For omega 3 fatty acid content QTL, three, two and one genes located within or in close proximity to QTL region identified on LG 11, 18 and 20 respectively. On LG 11 for QTL associated with EPA, two genes, peroxisomal *carnitine O-octanoyltransferase-like (crot)* located within the QTL region and *farnesyl pyrophosphate synthase-like (fdps)* located 600kb upstream of the QTL region, were up and down-regulated in the fast-growing fish liver, respectively. As for QTL associated with DHA on LG 11, one gene, *squalene epoxidase a (sqlea)*, located approximately 14kb upstream of the QTL region, was down-regulated in fast-growing fish. On LG 18 for QTL associated to ALA and EPA, two genes, *cytochrome P450-family 7-subfamily B-polypeptide 1-like (cyp7b1)* and *inositol monophosphatase 1-like (impa1)*, located within and 140kb upstream of the QTL region respectively and were down and up-regulated in fast-growing fish liver respectively. On LG 20 for QTL associated with ALA and EPA as well, one gene, *glutathione synthetase-like (gss)* was located approximately 23kb upstream of the QTL region and was down-regulated in the fast-growing fish liver. These results indicate the reliability of the QTL mapping using both microsatellite and SNP based linkage map as well as provide a high confidence for this study.

Table 3.16. Differentially expressed genes that lie within/close to the QTL region of associated traits of F₂ hybrid saline tilapia.

Traits	LG	Position of identified QTL (cM)	Associated markers (Flanking)		Differentially expressed genes (DEG)				
			Name of markers	Genetic position (cM)	Genomic position (bp)	Gene symbol and description	Genomic position (bp)	p-value*	Fold change (Fast vs slow)
180BW	2	16.1	66307	10.1	NIL (scaffold)	<i>sparc</i> , Secreted protein acidic and cysteine rich	465053-464161	0	2.11
			212092	30.6	386219				
			242870	46.8	14612903	FoxK1, Forkhead box protein K1-like	11140820-11138678	0	-2.59
180TL	6	46.5	242320	46.6	12965663				
						FoxK1, Forkhead box protein K1-like	11140820-11138678	0	-2.59
			242320	46.6	12965663				
180BT	7	81.3	OMO361	85.0	38380450-38381389	<i>smad3</i> , Smad family member 3-like	38064419-38061775	0	-2.72
			259590	77.3	32325128				
			OMO306	37.0	8989979-8991078	<i>usp38</i> , Ubiquitin specific peptidase 38	10482767-10481432	0	-8.00
C20:5n3 (EPA)	11	57.8	241939	32.4	11862605				
			85297	57.301	16846639	<i>crot</i> , Carnitine O-octanoyltransferase-like	16851102-16850496	0	2.25
			85311	57.805	16894308	(peroxisomal)			
C22:6n3 (DHA)	11	57.8	84826	60.255	15514531	<i>fdps</i> , Farnesyl pyrophosphate synthase-like	16149900-16149252	0	-48.33
			92943	49.781	7937936	<i>sqlea</i> , Squalene epoxidase a	7872891-7869064	0	-4.56
C18:3n3 (ALA)	18	72.1	92904	50.749	7854666				
			160147	69.215	26188866	<i>cyp7b1</i> , cytochrome P450-family 7-subfamily B-polypeptide 1-like	25673733 - 25684378	0	-5.00
			160070	70.272	26030795				
C20:5n3 (EPA)	18	72.1	98115	72.127	NIL(scaffold)	<i>impa1</i> , Inositol monophosphatase 1-like	26170118-26171402	0	7.68
C18:3n3 (ALA)	20	15.2							
			184395	15.178	19773314				
			183983	15.991	18321777	<i>gss</i> , Gluthathione synthetase-like	19796370-19802128	0	-3.27

*Bonferroni methodology of correction was applied and all differentially expressed genes were statistically significant ($p < 0.001$)

3.5.2 Tissue expression of DEGs in muscle and liver

The differential gene expression of DEGs that lie or are within close proximity of the identified QTL for growth and omega 3 content in the dorsal muscle and liver of salt tolerant tilapia were analyzed in triplicates using qPCR (Table 3.17). Out of the ten DEGs, four genes, *forkhead box protein K1 (foxK1)*, *ubiquitin specific peptidase 38 (usp38)*, *farnesyl pyrophosphate synthase-like (fdps)* and *glutathione synthetase-like (gss)* were significantly differentially expressed between the fast and slow growing muscle while only two out of four genes, *foxK1* and *cytochrome P450-family 7-subfamily B-polypeptide 1 (cyp7b1)* were significantly differentially expressed in the liver (Table 3.17).

Table 3.17. Average gene expression relative to actin genes analyzed by qPCR, standard deviations and significance of DEGs that lie within the QTL regions in dorsal muscle and liver of fast and slow growing F₂ hybrid salt tolerant tilapia.

Gene name (symbol)	Muscle (n=3)			Liver (n=3)		
	Slow	Fast	<i>p</i> -value	Slow	Fast	<i>p</i> -value
<i>sparc</i>	1.32 ± 0.19	2.07 ± 0.10	0.27	2.93 ± 3.52	57.15 ± 5.29	0.16
FoxK1	5.07 ± 0.69	1.51 ± 0.15	0.03	1.65 ± 0.48	16.60 ± 1.16	0.00
<i>smad3</i>	6.06 ± 0.36	1.78 ± 0.18	0.18	7.03 ± 1.10	13.03 ± 2.65	0.53
usp38	17.63 ± 0.09	1.88 ± 0.09	0.01	40.89 ± 5.36	1.89 ± 0.99	0.21
<i>crot</i>	3.40 ± 3.27	1.48 ± 0.13	0.13	14.78 ± 5.82	94.69 ± 17.19	0.33
fdps	1.62 ± 0.21	1.29 ± 0.06	0.01	15.06 ± 3.20	9.07 ± 1.57	0.70
<i>sqlea</i>	2.22 ± 0.94	1.35 ± 0.32	0.16	98.25 ± 14.90	9.06 ± 1.10	0.25
<i>cyp7b1</i>	1.75 ± 0.31	2.05 ± 0.30	0.67	16.94 ± 2.42	1.46 ± 0.91	0.00
<i>impa1</i>	N.a	N.a	-	2.33 ± 0.43	9.47 ± 1.41	0.11
gss	54.71 ± 15.71	1.85 ± 0.12	0.01	34.19 ± 3.62	36.63 ± 3.14	0.26

One-way ANOVA was conducted on differential gene expression between fast and slow growing fishes. Significant differential gene expression between the fast and slow growing fishes are in bold with *p*-value <0.05. N.a; Not amplified

Chapter 4. Discussion

4.1 Construction of first densest linkage map of saline tilapia using both microsatellites and SNP markers

DNA markers are an important component of linkage maps that are crucial to QTL mapping, serving as a guide in the identification of chromosomal regions and genes associated with the traits of interest. Microsatellites are among the most popular DNA marker since they are co-dominant, highly polymorphic, highly repeatable and they can be easily and accurately scored (Liu & Cordes 2004). The high heterozygosity of the microsatellites enables a dramatic decrease in the number of samples utilised for further downstream analyses such as linkage map and QTL mapping for traits of interest (Vignal *et al.* 2002). With the advancement of electrophoretic technologies, the genotyping and sizing of the microsatellite alleles could be achieved with ease and cost efficiency as compared to other DNA markers. This is mainly because of the utility of fluorescently labeled primers and capillary electrophoretic machines with suitable software for genotyping. In addition, microsatellites can be utilized in closely related species therefore, they have been extremely popular in fish species for many applications such as parentage assignment, analyses of population structure, product traceability (Yue 2014) and evolutionary studies (Sun *et al.* 2009) in the past decades.

In this Thesis, a genome scan for QTL for growth and sex was conducted in F₂ hybrid salt-tolerant tilapia using four linkage maps that were microsatellite-based (Section 3.2.1). Generally, the linkage maps with marker intervals of less than 20 cM and population size ranging from 100-250 individuals would suffice (Mohan *et al.* 1997). However, genome size and complexity of the organism can affect the quality of the linkage map (Collard *et al.* 2005; Gjedrem & Baranski 2009) and thus, the QTL mapping. The genome size of tilapia is at approximately 1 Gigabase (Gb) (Gregory 2001) which is about the third of the human genome. Increasing the population size and careful selection of DNA markers to evenly cover the tilapia genome may allow for a linkage

map with higher resolution. The microsatellite-based linkage map in my study (Table 3.6) is constructed using 522 individuals and is approximately 100 cM longer than the previously published map with a total map length of 1067.6 cM containing 401 markers (Liu *et al.* 2013a). This may be because the number of markers used in my study was smaller and may lead to bigger gaps between the markers as well as decrease the resolution of the linkage map. Therefore, increasing the density of markers through NGS technologies (i.e. GBS) would allow for higher resolution linkage maps (Davey *et al.* 2011) and improve the accuracy of QTL mapping as well as serve as a basis for positional cloning (Darvasi *et al.* 1993).

In this Thesis, I have also looked at recombination ratios between female and male linkage maps (Table 3.7). Understanding meiotic recombination is important in QTL mapping for economical traits in fishes where sex determination systems vary. This is because recombinant rates observed are able to provide information of the marker linkage to QTL and the sex with the lower recombination rate is expected to transmit marker to QTL associations in tighter linkage (Lenormand & Dutheil 2005). Significantly different recombinant ratios between the female and male linkage maps have been reported in several teleost species such as Atlantic salmon (Moen *et al.* 2004), rainbow trout (Sakamoto *et al.* 2000) and European seabass (Chistiakov *et al.* 2005). In my study, the recombination rate between female and male linkage maps is 1.0 and this ratio was inconsistent throughout all linkage groups agreeing with the previous study conducted in *Oreochromis spp.* where the recombination ratios were only significant in a few LGs (Liu *et al.* 2013a). Crossover events in males and females unevenly distributed through the genome may potentially indicate beneficial combinations of alleles and this is advantageous for the survival of the species (Lorch 2005). Previous studies have also reported that this inconsistency is observed in other species of fishes such as turbot (Bouza *et al.* 2007), Atlantic halibut (Reid *et al.* 2007) and flounder (Castaño-Sánchez *et al.* 2010). However, it could also be possible that the marker density could have caused an underestimation of recombination rates between females relative to males. Therefore,

increasing marker density to shorten the marker intervals would allow for a better examination of this issue.

The utility of bi-allelic single nucleotide polymorphisms (SNPs) as DNA markers has been increasing rapidly in many genetic studies in aquaculture (Liu & Cordes 2004). Compared to multi-allelic microsatellites, they are the preferred markers for many applications in genomics and genetics such as linkage mapping due to the increased density they can provide (Evans & Cardon 2004). The key advantage of SNP over microsatellite markers is their potential for rapid, low-cost genotyping due to the ability to be fully automated, allowing for lesser human based errors compared to microsatellites (Pompanon *et al.* 2005). One of the popular SNP discovery platforms for both model and non-model organisms is the restriction site associated DNA sequencing (RAD-seq; (Miller *et al.* 2007)). RAD-Seq procedures allow for the reduction of the genome complexity and generate multitudes of DNA markers based on SNP nearby or between restriction sites scattered throughout the genome (Davey *et al.* 2011; Peterson *et al.* 2012). Amongst the different applications, genotype by sequencing (GBS) is most frequently used in non-model organisms without reference genomes (Elshire *et al.* 2011; Poland & Rife 2012). This is because GBS has the potential to be very cost efficient through intensive multiplexing and genotype large amounts of samples (Poland & Rife 2012). However, it is highly dependable on the sequence read depth (Beissinger *et al.* 2013) and the quality of the reads. This is especially the case for species with large genomes, where appropriate use of restriction enzymes is crucial for the reduction of genome complexity and ensuring that there is sufficient overlap of reads (Elshire *et al.* 2011). In addition, the assignment of genotypes are prone to error due to adapters with high rates of synthesis errors and must be corrected to ensure the accuracy of data (Hamady *et al.* 2008). Even though GBS is a quick and high throughput method of discovering and genotyping SNPs, special care must be taken to ensure that assignments of genotypes are accurate and imputation of genotypes are close to real data (Marchini & Howie 2010). With the decreasing cost of NGS technologies, whole genome

sequencing studies have been conducted in more than 20 aquaculture species (Yue & Wang 2017). Through whole-genome sequencing, large numbers of SNPs have been identified for example, in Atlantic salmon (Yáñez *et al.* 2016) and channel catfish (Sun *et al.* 2014a), 9.7 and 8.4 million putative SNPs have been identified respectively.

In this Thesis, I have identified 37,141 SNP markers using the GBS technology, from which 927 SNPs were genotyped in the F₂ salt tolerant hybrid tilapia mapping family (Section 3.2.3.1). When comparing with previous tilapia linkage maps (Lee *et al.* 2005; Liu *et al.* 2013a), my study has constructed much denser linkage map consisting of 784 DNA markers; 151 microsatellite markers and 633 SNP markers (Table 3.8). Even though large numbers of SNP markers were identified, the construction of a good quality linkage map can only be achieved by maximizing the number of recombinations through an increment in population size. In my study, 356 F₂ individuals from one family were utilized, which is one of the largest progenies in tilapia and the second largest amongst cichlids (n=382) (Recknagel *et al.* 2013). The markers were selected to maximize the number of loci and ensure that there is a minimum amount of data lost from each individual so as to increase the accuracy of the marker distances which are important in linkage map construction as well as QTL analyses (Hackett & Broadfoot 2003).

During the recent years, advances in the discovery and genotyping of SNPs has led to a decrease in both discovery time and genotyping cost, leading to a large number of SNPs being identified. Efficient genotyping platforms have also been developed to genotype these SNPs through the utilization of microarray chips (Syvanen 2005), allowing for the genotyping of large numbers of individuals in an efficient and accurate manner, as well as for generating high-density linkage maps. Currently these high-density SNP arrays have only been developed in for a few fish species; Atlantic salmon (Houston *et al.* 2014; Yáñez *et al.* 2016), common carp (Xu *et al.* 2014a), channel catfish (Liu *et al.* 2014c) and rainbow trout (Palti *et al.* 2015a) due to the high cost involved (Yue & Wang 2017). The cost-effectiveness of such SNP array decreases due to its

customization, where it has to be designed to enable a wide usability across species and/strains and to cover rare SNPs, which may be present in small numbers (Thomson 2014). Additionally, ascertainment bias may be introduced when selecting informative SNP and may present a biased view of the genetic relationships rather than a set of random, neutral loci (Thomson 2014). It is this high developmental cost prevents the application of such arrays in most aquaculture species. Nonetheless, the decreasing cost of next-generation sequencing (NGS) technologies have made available large amounts of genomic and SNP data, and the number of these SNP arrays being developed in aquaculture species will gradually increase.

SNP arrays are an important element in genome-wide association studies (GWAS), where the DNA markers can be examined thoroughly in the genome in large numbers of individuals in association with a trait. The high accuracy and reproducibility have shaped GWAS to be a cost-efficient method for accelerating the genetic improvement of traits in many species (Hong *et al.* 2012). GWAS is an essential step towards genomic selection (GS) (Yue 2014), where a high density of markers covering the genome are used to investigate the linkage between the DNA marker and trait, as well as the effects of this linkage on the phenotypic variation in the population and/or species (Meuwissen *et al.* 2001; Yue & Wang 2017). To date, there has not been any GS carried out in aquaculture species and GWAS was only conducted in a few species for several traits such as for disease resistance in Asian seabass (Wang *et al.* 2017), gilthead seabream (Palaikostas *et al.* 2016) and channel catfish (Geng *et al.* 2015; Zhou *et al.* 2017), for sex determination in European seabass (Palaikostas *et al.* 2015) and for growth rate and sexual maturation in Atlantic salmon (Gutierrez *et al.* 2015) and head size in catfish (Geng *et al.* 2016). Therefore, the SNP discovered in my study together with the existing genomic data such as the 1.4 million SNP identified (Xia *et al.* 2014b) and genomic assembly of Nile tilapia (Conte *et al.* 2017), lays the foundation for the development of the SNP array, which will greatly increase the progression of the genetic improvement of traits in salt tolerant tilapia.

4.2 Improved growth performance of saline tilapia

Growth is one of the most important traits in aquaculture. It is quantitative trait where its variation is determined by action of alleles in different loci (Falconer & Mackay 1996; Gjedrem & Baranski 2009). Quantitative traits are measurable phenotypes that vary among individuals and produce a continuous distribution of phenotypes, giving rise to bell-shaped normal distribution curve as observed in the mapping family in this data (Figure 3.1). In tilapia, males tend to be bigger in size when compared to females (Lorenzen 2000). Such an association of growth and sex is not unusual; it has been described in many species where growth rates have been known to be correlated (Bhatta *et al.* 2012). Similarly, in my study, the males in this mapping family were approximately 40g heavier, 20 mm longer and 4mm thicker than females at 180 dph (Table 3.1). It is this sexual dimorphism in growth that improves productivity and the advantage of being able to prevent unwanted mating that gives rise to the preference of all male production in tilapia (Beardmore *et al.* 2001).

Fish growth is positively allometric and applies for body thickness or 'plumpness' (Froese 2006) and in my study, the moderate to high correlation between body weight, total length and body thickness (Table 3.2) are in agreement with this statement. Different species of tilapia differ in their growth performance. In my study, I have used an F₂ population of hybrid tilapia that were descendants of a cross between Mozambique and red tilapia of unknown genetic background. Mozambique tilapia have been known to have lower growth performance when compared to Nile tilapia and are only used for hybridization. On the contrary, red tilapias are popular due to their colour, salinity tolerance and growth performance (Gupta & Acosta 2004). Even though the genetic background and selection histories of red tilapia are complex, previous studies have shown that red tilapia hybrid such as *O. mossambicus-hornorum* hybrid X *O. niloticus* (Romana-Eguia & Eguia 1999) and the hybrid of *O. mossambicus* X *O. niloticus* (Pongthana *et al.* 2010) fare better in terms of growth performance in saline

environments. Additionally, a comparison of growth rates and salinity tolerance between Nile tilapia, Mozambique tilapia, and their hybrids have shown that hybrids had a high positive heterosis for weight (1.24) and biomass (1.33) gain outperforming their purebred parents (Kamal & Mair 2005).

In this Thesis, the body weight of the fishes in my mapping family is comparable to those of the GIFT (288.82 g) and Supreme (288.59 g) tilapias, but double those of the red tilapia (151.17 g) (Santos *et al.* 2013). The fact that they are all part of an all-male breeding programme that selects for quick growth, is indicative that my project is going towards a positive direction. In addition, another comparison with the previous growth performance of the hybrid tilapia cultured in freshwater in a reciprocal cross (Liu *et al.* 2014a), the individuals in this study that were cultured in seawater (30ppt) is almost double the weight, showing superior growth performance in saline environments. The growth data in my study suggest that hybridization of salt tolerant Mozambique tilapia with the genetically improved red tilapia would be a feasible way to generate fast growing salt tolerant tilapia. I am inclined to think that by combining conventional breeding with MAS/GS, it is highly possible to breed a new strain of tilapia, which could grow much quicker and adapt to full seawater (e.g. >30 ppt), thereby exploiting the vast resource of the ocean without compromising on the desired traits and productivity of culturing saline tilapia.

4.3 Increased neutral lipid content and Omega 3 fatty acid of saline tilapia

Marine and freshwater fishes require different essential fatty acids in their diet. This is because marine fishes have lost their ability to synthesize enzymes required for the production of highly unsaturated fatty acids (HUFA) and therefore require supplementations of HUFA in their diet (Sargent *et al.* 2003; Tocher 2003; Tocher & Glencross 2015). However, most freshwater fishes possess delta 5 desaturase and are

able to desaturate omega 3 and 6 polyunsaturated fatty acids (PUFA) (i.e. linoleic acid and alpha-linolenic acid) to synthesize HUFA (Nakamura & Nara 2004). This differential ability to synthesize enzymes required for the production of highly unsaturated fatty acids (HUFA) such as eicosapentaenoic acid, EPA (C20:5) and docosahexaenoic acid, DHA (C22:6) from the omega 3 family meant that obvious differences would be observed in the lipid content and composition between marine and freshwater fishes.

Freshwater tilapia has been classified as lean fish (<5% total fat content) (Rasoarahona *et al.* 2005) and reports studying fatty acid composition in freshwater tilapia have been generally unfavourable, listing it as a fish that is low in lipids in the muscle (Karapanagiotidis 2017). Research on the fatty acid composition of tilapia is usually conducted with fishes cultured in freshwater and data can be contradictory due to various factors such as the age and size of fish, feed formulation, different extraction protocols and etc. Different studies in the Nile tilapia showed variation in the total neutral lipid content.

In this Thesis, the measuring of the total neutral lipid content and profiling of fatty acids is carried out in triplicates using the dorsal muscle of 246 F₂ individuals that are cultured in seawater (Table 3.3). When comparing to the total neutral lipid (17.3 mg/g) extracted in the individuals in this study to tilapia cultured in freshwater, it was 50% higher than wild Nile tilapia (5.82 to 6.96 mg/g) (Karapanagiotidis *et al.* 2006) and sex reversed male Nile tilapia (7.03 to 7.82 mg/g) (Karapanagiotidis *et al.* 2007) and comparable to farmed Nile tilapia (11.83-21.84 mg/g) and red hybrid tilapia (12.63 to 18.01 mg/g) (Karapanagiotidis *et al.* 2006). In Mozambique tilapia and its hybrid, it was reported that the amount of total lipids was approximately 2 g/100 g, which is comparable to this study (Rahnan *et al.* 1995). However, the data in that study is not reliable as the samples were dorsal muscles with skin hence the fat content measured is 'contaminated'. Altogether this is contradictory to the

common notion that tilapias have low lipid content in their muscles as it can be observed from previous studies that hybrids and farmed tilapia can, in fact, be 'fatty'.

In my study, the content of palmitic acid (C16:0, 18.5%) and oleic acid (C18:1, 21.5%), the dominant SFA and MUFA respectively (Table 3.3) were comparable to freshwater Mozambique tilapia from different provinces from South Africa (C16:0; 19.6-24.07%, C18:1; 16.09-21.98%) (Chauke *et al.* 2008), but were lower than the highly selected freshwater Nile tilapia (GIFT), where C16:0 was 26.84% and C18:1 was 30.60% of the total amount of fatty acids identified (Nguyen *et al.* 2010c). This attributed to the positive benefits of the fishes in my study as decreased consumption of SFA such as palmitic acid has been known to decrease the risk of coronary heart disease (Zong *et al.* 2016). It has been reported that farmed tilapia contained lesser omega 3 fatty acids than omega 6 fatty acids, and the high ratio between omega 6 and omega 3 in tilapia is detrimental to human health (Young 2009). Interestingly, in my study, the percentage of omega 3 fatty acids ($25\% \pm 0.8\%$) (Table 3.3) was much higher than that of the Mozambique tilapia (6-8%) (Chauke *et al.* 2008) and Nile tilapia (GIFT) (Nguyen *et al.* 2010c), which were both cultured in freshwater. Rather, it was comparable to that of other fatty fishes such as chinook salmon (28.1%), mackerel (23.8%), menhaden (27.4 %) (Castell 1979) and Asian seabass (31.1%) (Xia *et al.* 2014a). It is a common perception that marine fishes contain a higher amount of HUFA (Li *et al.* 2011). In my study, salt tolerant tilapia has shown to be comparable in that respect. I speculate that this may possibly be due to the fact that innately, freshwater fishes possess the delta 5 desaturase, while marine fishes do not. Therefore, they can efficiently produce HUFA without extra supplementation to their diets (Nakamura & Nara 2004). Clinical and epidemiological studies have suggested that HUFA found in fishes are extremely beneficial for heart health in humans (Leaf & Weber 1988; Mozaffarian & Wu 2012).

In my study, high amounts of HUFA in salt tolerant tilapia suggests that salt tolerant tilapia confers more positive health benefits upon consumption when compared to other freshwater tilapias. In addition, the exploitation of natural sea resources for its culture would increase profitability and decrease the need for land resources.

In human diets, the ratio of omega 6 to omega 3 fatty acids are indicative of the quality of the fish and it has been recommended that this ratio be less than 1-4:1 (Simopoulos 1999). Freshwater tilapia has been reported to have high omega 6 to omega 3 ($n6/n3$) ratio, and consumption of freshwater tilapia may promote inflammation and aggravate heart health (Young 2009). In my study, the ratio of $n6/n3$ in salt tolerant tilapia is 0.9 and its inverse $n3/n6$ is 1.2. This is well within the recommended ratio, making saline tilapia an ideal fish for consumption. In my study, correlation studies between the individual fatty acids and between each fatty acid (Table 3.4) and growth-related traits (Table 3.5) have also been conducted. This enables for a global view on the changes that may occur should selection be practiced for one trait or another. Generally, the correlation between the fatty acids was synergistic and improvements in one group of fatty acid are likely to have a positive effect on the whole. This is similar to the correlations observed in Nile tilapia selected for quick growth (Nguyen *et al.* 2010c) as well as Asian seabass (Xia *et al.* 2014a). On the contrary, the correlation between individual fatty acids and growth-related traits (i.e. BW and TL) in my study were positive (Table 3.5) when compared to fast-growing Nile tilapia, where most of the individual fatty acids were negatively correlated (Nguyen *et al.* 2010c). This could be due differences in culture environment as the later was cultured in freshwater. Moreover, studies conducted in several species of fish such as rainbow trout (İbrahim Haliloğlu *et al.* 2004), Atlantic shad (Liu *et al.* 2017), steelhead trout (Sheridan *et al.* 1985) and masu salmon (Li & Yamada 1992) have showed that changes in salinity can affect fatty acid composition

in fishes. To date, few studies have been performed on the effect of growth selection on lipid composition in fishes except in Nile tilapia where there seems to be no significant effect (Nguyen *et al.* 2010c). On the other hand, studies in farmed animals such as rabbits (Pascual & Pla 2007) and pigs (Cameron *et al.* 2000) have shown that increasing growth rate resulted in favourable changes to the lipid composition. Therefore, future studies such as the development of more DNA markers (microsatellite, SNP and gene-based markers) for QTL mapping using larger sample sizes to conduct MAS/GWAS, or the fine-mapping and positional cloning of genes related to HUFA biosynthesis and growth, must be carried out to improve favourable HUFA content in the muscle of salt tolerant tilapia.

4.4 Increased growth traits in saline tilapia associated with multiple genomic regions

Growth is a polygenic trait and is regulated by both genetic and environmental factors (Reinecke 2010). Being one of the important economic traits, QTL mapping for growth-related traits has been conducted in many fish species such as Atlantic charr (Küttner *et al.* 2011), Atlantic salmon (Baranski *et al.* 2010; Tsai *et al.* 2015), Asian seabass (Wang *et al.* 2015a), European seabass (Louro *et al.* 2016) and rainbow trout (Kocmarek *et al.* 2015). Generally, most of the QTL were mapped within a small region of less than one million base pairs and candidate genes were speculated from these regions, but in order to identify causal genes, fine-mapping of these QTL regions would be necessary.

In this Thesis, the QTL for growth traits in all offspring using the microsatellite-based linkage maps showed rather small effects (Section 3.3.1.2) and they differ from those found in previous studies (Cnaani *et al.* 2003; Liu *et al.* 2014a). This reflects the nature of QTL for polygenic traits (i.e. growth) across

species and/or families within a species as different crosses as the recombination between the markers and QTL differs due to the contribution of the parental strain (Wu *et al.* 2007). Additionally, the difference in the sensitivity of QTL detection can be influenced by an effective population size, structure of the population, environmental effects, density of the markers in the linkage map, genotyping errors and accuracy during the measurement of the traits of interest (Paterson *et al.* 1988; Doerge 2002). In my study, QTL significantly associated with sex was detected in LG 1 and LG 22 (Figure 3.3 and Table 3.9) and previous studies have mapped the sex determination loci on LG 1 (Cnaani *et al.* 2008; Liu *et al.* 2013a; Palaiokostas *et al.* 2013), LG 3 (Cnaani *et al.* 2008) and LG 22/23 (Shirak *et al.* 2006; Eshel *et al.* 2011; Liu *et al.* 2013a). This is suggesting that the salt tolerant family in my study, the determination of sex may be associated with two different genomic locations. The same phenomena has been observed with the black-spotted stickleback where LG 12 and LG 19 were associated with sex (Ross *et al.* 2009), hence, further cytogenetic studies, as well as an increase in the number of markers within the QTL region, will shed better light of sex determination in the salt tolerant tilapia family in my study. As my study is more focused on the detection of QTL for growth, separate analysis of QTL for growth in male and female offspring is conducted due to the small effects of the QTL observed in all offspring (Section 3.3.1.2). This separation of sexes has increased the power of detecting QTL for growth traits in salt-tolerant as observed from the larger QTL effects (Table 3.9). In aquaculture production, tilapia males are preferred due to their better growth performance (Macintosh & Little 1995). Recently, markers have been developed specifically targeting MAS in male monosex culture of Nile tilapia (Sun *et al.* 2014b) but there was no association with growth performance in that project, therefore, the findings in my study are particularly encouraging. The significant QTL detected on LG 2 and

LG 18 in the male offspring corresponded to chromosome 4 and 3 in the stickleback (Guyon *et al.* 2012), which was reported to be associated with gain and loss of skeletal traits (Miller *et al.* 2014). This is especially the case with chromosome 4, which contained a trait cluster associated with both increase and decrease in skeletal morphological traits. Additionally, transferrin, an iron-binding glycoprotein that is known to play an important role in osmoregulation and salt tolerance (Rengmark & Lingaas 2007), was first mapped onto LG 21 (Cnaani *et al.* 2002) and more recently, onto LG 18 (Lee *et al.* 2005). Hence, it is possible that the gene(s) underlying the QTL regions in LG 2 and LG 18 may be tightly linked or have a pleiotropic effect on the multiple growth traits in male salt tolerant tilapia. Interestingly, the suggestive QTL detected on LG 15 in male offspring was previously reported to harbor the estrogen receptor (Guyon *et al.* 2012), which was found to show sexually dimorphic expression with higher transcript levels in males. On the other hand, in the female offspring, the significant QTL detected on LG 8 corresponded to chromosome 5 in the stickleback (Guyon *et al.* 2012) and was also reported to be associated to skeletal associated traits albeit different traits such as gain of branchial bones (Miller *et al.* 2014) and number of lateral plates (Liu *et al.* 2014b). In all, the QTL detected on these LGs suggests that pleiotropic effects may influence the growth of fishes in a sex-specific manner. QTL was detected on LG 22 for body weight and was also significantly and suggestively associated with sex and TL in all offspring, respectively. The QTL was also associated with both BW and TL in the male offspring. On the contrary, QTL detected on LG 1 was only significantly associated with sex but was suggestively associated with BW and TL in female offspring (Table 3.9). Previous studies have reported LG 1 and LG 22/23 to be associated with sex determination (Cnaani *et al.* 2008; Liu *et al.* 2013a; Palaiokostas *et al.* 2013; Ma *et al.* 2016a) hence, it is suggestive of the gene interaction that is underlying the sexual

dimorphism of growth in salt tolerant tilapia in my study. Besides the QTL detected on LG 22, another two overlapping QTL of genome-wide and chromosome-wide significance for growth traits in LG 8 and LG17 were detected in both male and female offspring as well. Even though most of the suggestive QTL detected in my population was not studied in detail, the differing location and overlaps of the QTL for sex and growth traits in male and female offspring suggest that different sets of genetic 'switches' are at play. Previous studies of QTL mapping for growth traits in hybrid tilapia (Liu *et al.* 2014a) and Atlantic salmon (Tsai *et al.* 2015) and early maturation in rainbow trout (Haidle *et al.* 2008) have shown that separate QTL analyses could yield more information to discover key overlapping regions associated with the trait of interest. Therefore, in my study, the key overlaps gave us further information and further analyses such as fine-mapping these regions and functional characterization of genes residing in these regions should be carried out, as this may allow for a better understanding of the phenotypic variance associated with the trait of interest in different sexes as well as population-wide.

Given the polygenic nature of growth traits (Paterson *et al.* 1988), it is not surprising that there could be many QTL regions of small effects that are also associated with other traits. Hence, increasing the resolution of the linkage map may allow for an increase in the genetic resolution of these overlapping QTL and a better understanding of the effect epistasis on conferring the growth variance in the population. In order to increase the density of the linkage map for fine QTL mapping of growth traits, I have utilized GBS technology and constructed a much denser linkage map.

Using this linkage map, QTL mapping for growth traits in F₂ offspring was carried out and three linkage groups were associated with BW, TL and BT significantly and suggestively (Table 3.10). When comparing the QTL mapping

results of all offspring using microsatellite-based linkage map, it is obvious that the QTL detected are of bigger effects and of a different location and more QTL with smaller effects were detected. These results affirm that by increasing the density of the linkage map, the power of QTL detection could be increased. A comparison of the newer and denser linkage map with all male and all female analyses, similarities in the LG where QTL is detected is observed. For example, LG 2 and LG 6 were significantly associated with BW and suggestively associated with TL and BT (Table 3.10) in the denser linkage map, while the QTL detected at LG 2 and LG 6 was previously associated significantly with BW and TL of male offspring and suggestively with BW, TL and BT of female offspring (Table 3.9). Another example is LG 17, where it was shown that one of the overlapping QTL that was suggestively associated with BW and TL in male offspring and TL in female offspring (Table 3.9), is now only suggestively associated with TL (Table 3.10) when using the denser linkage map. As mentioned previously, LG 2 corresponded to chromosome 4 of the stickleback and similarly, LG 17 corresponded to the same chromosome 4. This association with multiple skeletal traits (number of lateral plates, raker, teeth, branchial, jaw, fin and vertebrae) in the chromosome 4 of the stickleback (Liu *et al.* 2014b; Miller *et al.* 2014) is indicative of the significance of LG 2 and 17 with growth-related traits in saline tilapia. Additionally, LG 6 corresponded to chromosome 9 in the stickleback, which is similarly associated with multiple skeletal traits (number of lateral plates, fin, teeth and jaw development in stickleback) as well, explaining the phenotypic variance of 9.1% (Liu *et al.* 2014b) and 4.4% to 7.0% (Miller *et al.* 2014). Even though in my study, suggestive QTL on LG 6 detected in this study had only explained low moderate phenotypic variance from 3.6% to 5.2%, it is still a genomic region that I have to factor in, where there may be a positive implication to the phenotypic variation that was observed in my population

of salt tolerant tilapia. Future studies to increase the marker density and therefore increasing the resolution of the linkage map to less than 1cM marker intervals should be conducted, in order to increase the power for detecting QTL for important traits. This can be done through fine-mapping of the QTL regions in LG 2, 6 and 17 and identification and positional cloning of candidate genes and the effects of epistasis on the growth variance in the population. Consequently, future efforts will be made to conduct QTL analyses for growth in more salt tolerant tilapia families cultured in full seawater to identify more growth QTL regions. Using multi-family mapping approaches will allow for the evaluation of previously identified QTL as well as increase the resolution of the QTL position and accuracy of its effects (Sneller *et al.* 2009). Nonetheless, all the markers associated with the detected QTL regions could be included into a repertoire of markers that are utilized for the improvement of growth of saline tilapia.

4.5 Improved omega 3 fatty acid content in saline tilapia is associated with four genomic regions

There is a high consumer demand for fishes with high fat content for flesh quality and omega 3 and 6 fatty acids (Hearn *et al.* 1987) as their consumption is beneficial to human health (Sidhu 2003; Weichselbaum *et al.* 2013). Lipids are crucial for normal growth and development of fishes (Sargent *et al.* 2003) and can be influenced by diet, growth rate (Fauconneau *et al.* 1995) and environmental (Whatmore *et al.* 2013) and genetic factors (Fauconneau *et al.* 1995) making the selection of fishes with high lipid content and high omega 3 content in the fish flesh much more complex. Even though traditional breeding methods are noteworthy for traits that are easily observable such as growth (Gjedrem & Baranski 2009), the measurement of fat content and lipid profile of the fish brings about a challenge due to its terminal sampling, high labour cost and time required to phenotype. So far, there has been slow progress in the studies of the genetic basis of muscle fat

content and lipid composition in fishes. In a few species, heritability was measured as part of a breeding programme and multiple families have been used to conduct such studies in an attempt to improve the fat composition of the fishes. For example, in common carp (Kocour *et al.* 2007) and yellowtail kingfish (Whatmore *et al.* 2013) the heritability for fat content was 0.58 and 0.41 respectively, in Atlantic salmon and rainbow trout, the heritability estimates vary widely from 0.19 to 0.38 (Rye & Gjerde 1996; Quinton *et al.* 2005; Vieira *et al.* 2007; Powell *et al.* 2008) and 0.25 to 0.40 (Quillet *et al.* 2005; Tobin *et al.* 2006) respectively depending on the strain. Even though the varying heritability observed in different species, it is indicative of the genetic basis of this trait. As there has been slow progress in mapping QTL for flesh quality and lipid traits in commercial fishes, any form of comparison or referencing is largely limited.

In this Thesis, QTL mapping was conducted using F₂ salt tolerant tilapia and the huge variation in the individual fatty acids (Table 3.3) and the positive correlation of each fatty acid to growth-related traits; BW and TL (Table 3.5) was observed suggesting the plausibility of QTL analysis. Altogether, more than 30 QTL both significantly and suggestively associated with lipid content and composition were detected (Table 3.11). The high number of suggestive QTL of small effects were detected was lower than that of the Asian seabass (Xia *et al.* 2014a), presumably due to a higher resolution in the linkage map used in my study. Previous studies conducted in common carp (Kuang *et al.* 2015) and Atlantic salmon (Derayat *et al.* 2007) did not detect as many QTL due to the fact that only total fat content was measured and analysed and individual fatty acids and their associations with different chromosomal regions were unknown. In my study, the QTL associated with individual fatty acids overlapped each other when comparing within groups and between groups (SFA, MUFA, HUFA) (Table 3.11). This is not surprising since SFA and MUFA are precursors in the biosynthesis of HUFA (omega 3 and omega 6). Common genes such as elongases and desaturases in the pathway will show pleiotropic QTL effects on these traits. For example, QTL detected for C18:0 was suggestively and significantly associated chromosomally with LG 11 and 20 respectively,

and did not overlap any of the QTL detected in SFA and MUFA, rather it was on the within close proximity of approximately 1cM each of the QTL detected for DHA on LG 11 and four HUFA (ALA, C20:4n6, EPA, C22:4n6) on LG 20. Additionally, strong positive correlations have been observed between C18:0 and the above-mentioned HUFA (Table 3.4), suggesting that these chromosomal regions may contain gene(s) that are responsible for multiple roles in the biosynthesis of HUFA. Additional markers flanking these regions should be genotyped and will be able to provide more information to narrow down the QTL region. Another QTL region on LG 16 was also associated with four traits (TFA, C16:0, C18:1 and C20:3n6) with moderate to strong positive correlations between them, further indicating an interplay of the gene(s) that regulate the pathway, and thus warranting further investigation. Similar data have been reported in Asian seabass (Xia *et al.* 2014a) as well as in farmed animals such as pigs (Sanchez *et al.* 2007) and sheep (Karamichou *et al.* 2006) where QTL regions have been associated with two or more fatty acids.

For omega 3 fatty acids, significant and suggestive QTL were detected in the same location at four linkage groups, LG 11, 18, 19 and 20 (Figure 3.7 to Figure 3.10) chromosomally with different effects and there was a positive correlation between these omega 3 fatty acids and growth-related traits. These overlapping QTL regions and their association with multiple traits increase the confidence that indeed it may be possible to select for a fast-growing saline tilapia with good lipid composition. However, further studies through fine mapping and positional cloning of desaturases, elongases and fatty acid related genes such as fatty acid binding proteins (*fabp*), fatty acid synthases (*fas*) and peroxisome proliferator-activated receptors (*ppar*) and verifying these genes with other populations are required. Despite the fact that further studies need to be conducted, existing data of the markers associated with the detected QTL especially those overlapping regions associated with multiple traits can be utilized in MAS for the genetic improvement of fatty acid composition, namely the omega 3 and 6 fatty acid content in salt tolerant tilapia. This can be conducted through the verification of these DNA markers

and their association with the omega 3 and 6 content of different families of fast-growing salt tolerant tilapia followed by the identification of the beneficial combination of alleles that showed a significant increase in omega 3 fatty acid content in fast-growing salt tolerant tilapia.

4.6 Increased growth and improved lipid contents are associated with Hippo pathway, decrease in apoptosis and lipogenesis in fast-growing fishes

Growth in fishes has a significant impact on aquaculture production due to declining wild stocks and increasing demand for consumption (Naylor *et al.* 2000). The increase in consumption is mainly due to the benefits of omega 3 and 6 fatty acids on the human's health, thus, many studies have been focused on the improvement of growth and lipid content especially fatty acid content either through quantitative genetics and/or next-generation sequencing technologies (Liu 2007; Yue 2014; Sun *et al.* 2016). In fishes, the muscle and liver are two tissues that play a crucial role in the development and energy metabolism of the fishes (Glencross 2009). Despite the fact that both of the tissues behave differently in their own intrinsic ways, they are linked together as they are main sources of energy reserves in fishes to regulate growth and development of the fish (Jobling 2007).

In this Thesis, the focus was on the identification of candidate genes that may improve our understanding of the molecular mechanisms underlying growth and lipid biosynthesis by comparing the differentially expressed genes in fast and slow growing fish cultured in seawater. A reference transcriptome of the F₂ hybrid salt tolerant tilapia was assembled based on the possibility that there may be transcripts that are lineage-specific to either Red or Mozambique tilapia (Table 3.13). The number of transcripts (262,282) and average contig size (846 bp) were fairly close to that of the Asian seabass (267,616 contigs, 979 bp) assembled using multiple platforms (Thevasagayam *et al.*

2015) and blunt snout bream (253,439 contigs, 998 bp) (Tran *et al.* 2015). In contrast, number of contigs were much higher than those in the Mozambique tilapia (196,178) (Lam *et al.* 2014) and the red tilapia (160,762) (Zhu *et al.* 2016) while the average contig length of 846 bp was higher than that of the Mozambique tilapia (645bp) (Lam *et al.* 2014) and lower than that of the red tilapia (1120.61 bp) (Zhu *et al.* 2016). This difference in the quality of assembly is largely due to the difference in sampling, sequencing platforms and assembly algorithms and the reference transcriptome in my study is of a reasonable quality since the actual lineage of my strain is complex. Global gene expression and variation in transcription activity occur across the genome and the number and level of transcript isoforms are not always known (Costa *et al.* 2010; Oszolak & Milos 2011; Pelechano *et al.* 2013). Therefore, in my study, annotation was carried out using nucleotide and protein databases (BLASTN and BLASTX), bringing the number of annotated transcripts to 42,699, which is more than 70% of the total number of mRNA transcripts in the NCBI database for both Nile and Mozambique tilapia. This indicates that my reference transcriptome is comprehensive to a reasonable degree to represent different levels of expression abundance.

In order to identify candidate genes for growth and lipid biosynthesis, differential gene expression between fast and slow-growing fishes was conducted (Section 3.4.3) and the expression of DEGs was verified against the RNA sequencing data (Table 3.15). Even though there was a large number of differentially expressed transcripts in the muscle (2,236) and liver (3,020), the majority of the transcripts were down-regulated in both tissues in the fast-growing fish (Figure 3.12). More than 42.7% and 57.2% of the differentially expressed transcripts in the muscle and liver respectively were assigned GO terms (section 3.4.4), further indicating the good quality of the annotation of my reference transcriptome. Additionally, the DEGs identified for both muscle and liver (Figure 3.13) had diverse cellular functions and were mainly in pathways of metabolic processes, suggesting the complex crosstalk between the muscle

and liver in the regulation of metabolic processes for the growth and development of the fish.

Pathway analysis using KEGG allows for a global view on the possible molecular mechanism where these DEGs may play a role in conferring superior growth and higher meat quality in terms of fatty acid content (Figure 3.14). Previous tilapia transcriptome studies were not conducted for the identification of growth or lipid biosynthesis related but rather on effects of temperature (Hu *et al.* 2016b) and diet (Xiong *et al.* 2014), skin colour (Zhu *et al.* 2016), salinity (Lam *et al.* 2014; Xu *et al.* 2015) and sex (Wei *et al.* 2016). In my study, investigating several pathways associated with growth and lipid biosynthesis will be attempted. From the summarized on muscle DEGs and the roles they play in molecular pathways, several pathways will be briefly discussed due to the complexity of the crosstalk between signalling pathways (Appendix 11).

The Hippo pathway, which has been reported to promote cell death and differentiation and inhibit cell proliferation (Zhao *et al.* 2011; Yu & Guan 2013), showed decreased expression in fast-growing fish in my study. As fishes have indeterminate growth, this finding may be suggestive of a positive transcriptional regulation of cell growth in fast-growing fishes in my study. However, the crosstalk of the Hippo pathway with other signalling pathways warrants further investigation on how its down-regulation may confer superior growth in my fast growing fishes. Surprisingly, the mTOR pathway, a positive regulator of growth known to promote anabolic processes, including biosynthesis of proteins, lipids and organelles and limiting catabolic processes such as autophagy (Laplante & Sabatini 2009) was generally down-regulated in my fast growing fishes. This is in contrast with findings from studies conducted in rainbow trout where slow growing fishes had suppressed mTOR signaling (Danzmann *et al.* 2016). Nonetheless, genes of the mTOR pathway have suggested a more positive outlook since S6 kinase alpha-6 (*S6K*) which has been reported to play a positive role in programmed cell death (Schmid *et al.* 2008) is down-regulated. Another gene, eukaryotic translation

initiation factor 4E (*eIF4e*), which is reported to be part of the protein translation initiation complex (Sonenberg *et al.* 1979), is up-regulated as well. Additionally, increased gene expression of the calcium and muscle contraction pathway and decrease in the expression of caspase 3 (*cas3*), serine/threonine-protein phosphatase 2A (*ppcb2a*), growth arrest and DNA damage-inducible protein (Gadd45), forkhead protein O3 (*foxo3*), smad family member 3 and 4 (*smad 3/4*) has been observed. Altogether, this may be suggestive of enhanced protein levels in the muscle and that growth is being promoted in fast-growing fishes through the decrease of apoptotic signals. Further studies to ascertain the total protein levels of the dorsal muscle of fast-growing fishes and correlating them with the thickness of muscle fibers needs to be conducted to verify these assumptions.

For the differentially expressed genes (DEG) in the liver (Appendix 12), the general trend for fatty acid (FA) content associated pathways such as the biosynthesis of unsaturated FA, FA elongation, alpha-linolenic acid and linoleic acid metabolism is up-regulated in the fast-growing fishes. This is in contrast to salinity challenged tilapia, where, it was reported that osmotic stress decreases these FA pathways in order to cope with stress (Xu *et al.* 2015). In addition, the main catalyst of FA degradation, carnitine O-palmitoyltransferase 2 (*cpt2*), which was previously found to be affected by dietary changes in Atlantic salmon and predominates in the red muscle (Frøyland *et al.* 1998), was down-regulated slightly. To date, there have not been too many studies on the function of *cpt2* in lipid biosynthesis in the liver in fishes, however, accelerated growth has been reported in European seabass (Santulli & D'Amelio 1986) and African catfish (Torreele *et al.* 2007) that have been fed with a diet supplemented with carnitine. Genes in the peroxisome proliferators-activated receptors (PPAR) signalling pathway such as fatty acid desaturase 2 (*fads2*), retinoic acid receptor (*rxr*) and fatty acid binding proteins (Fabp) known to play a major role in lipogenesis and the biosynthesis of fatty acids (Nakamura *et al.* 2004), are generally up-regulated in my study. I think that this could

most probably explain the higher fatty acid content in the fishes in my study since lipogenesis and biosynthesis of fatty acids are up-regulated.

Compared to mammals, fishes have lower capabilities to use carbohydrates as an energy source but rather lipids are their main source of energy (Tocher 2003). Interestingly, in this study, the general trend of gene expression in AMPK signalling pathways in the muscle and liver contrasted each other. AMPK signalling pathway is known as the master regulator of cellular metabolism (Hardie 2004; Witczak *et al.* 2008) and studies in brown and rainbow trout have shown that that AMPK signalling and its genes such as *PGC 1-alpha* are involved in the energy metabolism and glucose uptake in the skeletal muscle (Magnoni *et al.* 2012) as well as up-regulation of fatty acid synthase (*fasn*) and acetyl-CoA carboxylase (*acac*) (Polakof *et al.* 2011). It is premature to say that the trends that have been observed in the differential expression of genes between the fast and slow-growing fishes in my study are positively correlative due to the complex crosstalk of the signalling pathways, however, it does lay the groundwork and direction for future studies in the functional characterization of genes related to growth and lipid biosynthesis in saline tilapia.

Apart from the complex crosstalk of signalling pathways, the main purpose of the identification of DEGs in the muscle and liver between fast and slow growing fishes was to allow for the discovery of candidate genes to better understand the phenotypic variation of growth and fatty acid content (omega 3) in salt tolerant tilapia. In this Thesis, ten DEGs from the muscle and liver were found to be located inside or within close proximity of QTL regions for growth and fatty acid content (Table 3.16) and correlated to the RNA sequencing data (Table 3.15). Surprisingly, DEGs from the liver (*fdps*, *gss*) that are associated to lipid biosynthesis was significantly expressed in the muscle and a DEG from the muscle (*foxK1*) related to growth was significantly expressed in the liver. On the contrary, the remaining DEGs that were differentially expressed in either muscle or liver pertaining to their association with growth or omega 3 fatty acids, respectively, have not shown significant in their difference in their expression level. In particular, the

foxK1 gene was significantly differentially expressed in both muscle and liver. Studies in mice have reported that *foxK1* promotes cell proliferation through (i) translational repression of forkhead box protein O4 (*foxo4*); (ii) inhibiting myocyte differentiation through myocyte enhancer factor 2 (*Mef2*) (Shi *et al.* 2012); (iii) repressed autophagy of muscle cells through the restriction of acetylation of histone H4; and (iv) expression of critical autophagy genes (Bowman *et al.* 2014). In addition, studies in tiger pufferfish have shown that there were three variants of the FoxK1 protein (alpha, gamma and delta) which were differentially expressed in many tissues, including the muscle and liver (Fernandes *et al.* 2007). It is interesting to note that the expression of *foxK1-alpha* agreed with the differential gene expression in the muscle between fast and slow-growing fishes in my study. It has also been reported in mice that *foxK1* plays a role in the regulation of natural killer (NK) cells and deficiency of *foxK1* causes a significant drop of NK cells in the spleen, blood and bone marrow. Additionally, NK cells are part of the innate immunity and play a role in liver inflammation in mammals (Gao *et al.* 2009). Nonetheless, the role *foxK1* plays in fish is still unknown hence, further studies on the role the role of *foxK1* in the liver of fishes pertaining to lipid biosynthesis would be interesting since it was observed to be up-regulated in the liver of fast-growing fish in my study.

Future studies would have to be conducted to understand the role of these DEGs especially those that are located within or in close proximity to detected QTL regions for both growth and fatty acid content. Future additional studies involving protein expression levels and/or enzymatic activity of the identified candidate genes could also be carried out to give us a more accurate representation of the role of these DEGs in the improvement of growth and increment of omega 3 fatty acid content in the fishes. For example, studying gene regulation using micro RNAs (miRNAs) such as miR-122, miR-33, miR-17, which have been reported to be regulating genes associated with lipid metabolism and HUFA in teleost (Zhang *et al.* 2014; Mennigen 2016) and miR-1, miR-133a and miR-206 that has been reported to be associated with skeletal growth in both

Nile tilapia (Huang *et al.* 2012; Yan *et al.* 2012a) and common carp (Yan *et al.* 2012c). Investigating the relationship between the differential gene expression of these miRNAs and the respective DEGs that was identified in my study, in fast and slow-growing, as well as higher fat and lower fat content individuals, might shed some light in the gene regulatory pathways in the growth and biosynthesis of HUFA with particular attention on omega 3 fatty acids. Another example is the utility of genome editing tools such as CRISPR-Cas9 technology, where precise changes such as insertion, deletion, and sequence changes can be made to the DNA (Sander & Joung 2014), for the investigate the role of these DEGs in conferring phenotypic variation in the salt tolerant tilapia. However, special care must be taken to ensure such genome editing to these genes do not cause lethality but rather modulate the gene expression, so that phenotyping can be conducted to verify the role of these DEGs in growth and HUFA in salt tolerant tilapia. The acceleration of the genetic improvement of growth and fatty acid content of salt tolerant tilapia can be achieved through the understanding of the roles these DEGs play in conferring phenotypic variance in the population.

4.7 Summary and future directions

Growth and omega 3 fatty acid content are important traits in tilapia breeding and understanding the molecular mechanisms underlying growth and omega-3 can accelerate the genetic improvement of these two important traits in salt tolerant tilapia. In this Thesis, I have constructed four linkage maps with 123-125 microsatellites. To increase the density of the linkage maps, I genotyped 356 individuals using GBS and I constructed a linkage map with 151 microsatellites and 633 SNPs. To the best of my knowledge, this linkage map is the densest map for saline tilapia. Using the two linkage maps, I have performed QTL mapping for growth and fatty acid content using both microsatellite-based and a combination of microsatellite and SNP based markers in this thesis. A mapping population of F₂ hybrid salt tolerant tilapia cultured in salinity of 30 ppt was genotyped, sexed and trait and tissue samples were collected at different ages (i.e. 110 dph, 180 dph and 220 dph). Four linkage maps for all offspring, only male, only female and male and female offspring consisting of 125, 124, 123 and 125 markers respectively covering 22 linkage groups were constructed. QTL analysis was conducted using multiple QTL mapping (MQM). Two QTL significantly associated with sex were identified in LG 1 and LG 22, explaining 19.8% and 10.8% phenotypic variance respectively. In addition, two and one QTL for growth explaining 6.3% to 10.1% and 8.5% to 10.2% of phenotypic variance in LG 2 and LG 18 respectively and 10.4% to 19.8% phenotypic variance on LG 8 was identified in male and female offspring respectively. These results suggest that besides the complex sex determination system in different species of tilapia, different sets of genetic 'switches' are at play in conferring growth performance. This agrees with the statement that growth is a complex and polygenic trait.

Constructing linkage maps with microsatellites limits identification of possible genes underlying the QTL due to its large confidence interval. The increase in the resolution of the linkage map through the increment of the number of markers will allow

for the identification of causative genes with high confidence. Using the denser linkage map, two significant QTL for body weight on LG 2 and LG 6 explaining 5.1% and 5.6 % of the phenotypic variance was detected. In addition, one QTL of genome-wide significance and 14 QTL of chromosome-wide significance in LG 8 and six LG were detected respectively. The QTL of genome-wide significance explained 8.3% phenotypic variation of fatty acid content in the population while the 14 QTL explained 5.9% to 17% phenotypic variance. The increment of omega 3 fatty acid content; C18:3n3 (ALA), C20:5n3(EPA) and C22:6n3 (DHA) in fishes has been a focus for QTL mapping of fatty acid content in my study due to the positive health benefits it brings. QTL located in four linkage groups have been detected to be significantly and suggestively associated with all three fatty acids, explaining 5.0% to 7.5% of phenotypic variation in the omega 3 fatty acid content. Markers linked to QTL for growth traits and fatty acid content can be used in MAS to accelerate the genetic improvement of growth traits and omega-3 contents.

In an attempt to identify candidate genes for growth and omega 3 fatty acid content, differentially expressed genes of the muscle and liver between fast and slow growing fishes were investigated using a *de novo* assembled reference transcriptome and pathway analysis with KEGG was conducted. Data suggests that fast-growing fish muscles may be promoted through the depression of the Hippo pathway and a decrease in apoptotic signals, and lipid biosynthesis is generally up-regulated in the fast-growing fish liver, correlating with the higher fatty acid composition in my salt tolerant tilapia when compared to other freshwater fishes. In addition, ten DEGs in the muscle and liver between fast and slow growing fishes were found to be located in the detected QTL regions for growth and omega 3 fatty acid content. Expression of these ten genes in the muscle and liver of fast and slow growing fishes showed correlation to the differential expression of these genes in the RNA sequencing data.

The QTL and candidate genes identified in this study lay the ground work to understand the molecular mechanisms and underlying cause of the phenotypic variation

of growth and omega 3 fatty acid content in salt tolerant tilapia. In addition, markers tightly linked to the QTL or polymorphisms in the candidate genes could be potential markers that could be implemented in the MAS programmes for improving growth performance and meat quality of salt tolerant tilapia. Due to the fact that differentially expressed genes can be found inside or within close proximity of the QTL regions and the initial pathway analysis conducted to identify candidate genes, further investigation through functional characterization of these candidate genes such as using CRISPR-Cas9 or TALEN genome-editing systems should be conducted.

The complexity of the mechanisms that underlie growth and fatty acid content cannot be undermined and further improvements to the density of linkage map through increasing the population size, number of families used in GBS will allow for fine-mapping of QTL regions with confidence. In addition, the large number of SNP markers discovered through GBS will allow for a more accurate investigation of the phenotypic variation occurring in the population through genome-wide association studies (GWAS) as they cover the whole genome and requires no pedigree information (Yue 2014). The identified SNPs from the GBS technology can be utilised in the construction of SNP microarrays, known as SNP chips, for the identification of SNPs associated with economically important traits (Yue & Wang 2017). Currently, the SNP chips have only been constructed in four food fish species due to the high cost of development. However, with decreasing cost of NGS technologies and increasing genomic data publicly available for tilapia, the construction of such SNP chips for the betterment of the growth performance and fatty acid content of the salt-tolerant tilapia is within reach.

References

- Acosta B.O. & Gupta M.V. (2010) The genetic improvement of farmed tilapias project: Impact and lessons learned. In: *Success Stories in Asian Aquaculture* (eds. by De Silva SS & Davy FB), pp. 149-71. Springer Netherlands, Dordrecht.
- Agaba M., Tocher D.R., Dickson C.A., Dick J.R. & Teale A.J. (2004) Zebrafish cDNA encoding multifunctional fatty acid elongase involved in production of eicosapentaenoic (20:5n-3) and docosahexaenoic (22:6n-3) acids. *Marine Biotechnology* **6**, 251-61.
- Agaba M.K., Tocher D.R., Zheng X., Dickson C.A., Dick J.R. & Teale A.J. (2005) Cloning and functional characterisation of polyunsaturated fatty acid elongases of marine and freshwater teleost fish. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **142**, 342-52.
- Ali M., Nicieza A. & Wootton R.J. (2003) Compensatory growth in fishes: a response to growth depression. *Fish and Fisheries* **4**, 147-90.
- Aljanabi S.M. & Martinez I. (1997) Universal and rapid salt-extraction of high quality genomic DNA for PCR-based techniques. *Nucleic Acids Research* **25**, 4692-3.
- Alves Martins D., Rocha F., Martínez-Rodríguez G., Bell G., Morais S., Castanheira F., Bandarra N., Coutinho J., Yúfera M. & Conceição L.E.C. (2012) Teleost fish larvae adapt to dietary arachidonic acid supply through modulation of the expression of lipid metabolism and stress response genes. *British Journal of Nutrition* **108**, 864-74.
- Arendt J.D. (1997) Adaptive intrinsic growth rates: An integration across taxa. *The Quarterly Review of Biology* **72**, 149-77.
- Arts M.T., Ackman R.G. & Holub B.J. (2001) "Essential fatty acids" in aquatic ecosystems: a crucial link between diet and human health and evolution. *Canadian Journal of Fisheries and Aquatic Sciences* **58**, 122-37.
- Arts M.T. & Kohler C.C. (2009) Health and condition in fish: the influence of lipids on membrane competency and immune response. In: *Lipids in Aquatic Ecosystems* (eds. by Kainz M, Brett MT & Arts MT), pp. 237-56. Springer New York, New York, NY.
- Atwood H.L., Tomasso J.R., Webb K. & Gatlin D.M. (2003) Low-temperature tolerance of Nile tilapia, *Oreochromis niloticus*: effects of environmental and dietary factors. *Aquaculture Research* **34**, 241-51.
- Avarre J.-C., Guinand B., Dugué R., Cosson J., Legendre M., Panfili J. & Durand J.-D. (2014) Plasticity of gene expression according to salinity in the testis of broodstock and F1 black-chinned tilapia, *Sarotherodon melanotheron heudelotii*. *PeerJ* **2**, e702.
- Avella M., Berhaut J. & Bornancin M. (1993) Salinity tolerance of two tropical fishes, *Oreochromis aureus* and *O. niloticus*. I. Biochemical and morphological changes in the gill epithelium. *Journal of Fish Biology* **42**, 243-54.
- Badyaev A.V. (2002) Growing apart: an ontogenetic perspective on the evolution of sexual size dimorphism. *Trends in Ecology & Evolution* **17**, 369-78.
- Bai Z.-Y., Han X.-K., Liu X.-J., Li Q.-Q. & Li J.-L. (2016) Construction of a high-density genetic map and QTL mapping for pearl quality-related traits in *Hyriopsis cumingii*. *Scientific Reports* **6**, 32608.
- Baranski M., Loughnan S., Austin C.M. & Robinson N. (2006) A microsatellite linkage map of the blacklip abalone, *Haliotis rubra*. *Animal Genetics* **37**, 563-70.
- Baranski M., Moen T. & Våge D.I. (2010) Mapping of quantitative trait loci for flesh colour and growth traits in Atlantic salmon (*Salmo salar*). *Genetics Selection Evolution* **42**, 17.
- Barrett L.W., Fletcher S. & Wilton S.D. (2012) Regulation of eukaryotic gene expression by the untranslated gene regions and other non-coding elements. *Cellular and Molecular Life Sciences* **69**, 3613-34.
- Bartel B. (2005) MicroRNAs directing siRNA biogenesis. *Nature Structural & Molecular Biology* **12**, 569-71.
- Beardmore J.A., Mair G.C. & Lewis R.I. (2001) Monosex male production in finfish as exemplified by tilapia: applications, problems, and prospects. *Aquaculture* **197**, 283-301.
- Beaumont A., Boudry P. & Hoare K. (2010) Genetic variation of traits. In: *Biotechnology and Genetics in Fisheries and Aquaculture* (pp. 99-128. Wiley-Blackwell.
- Beavis W.D. (1998) QTL analyses: power, precision, and accuracy. *Molecular dissection of complex traits* **1998**, 145-62.
- Beissinger T.M., Hirsch C.N., Sekhon R.S., Foerster J.M., Johnson J.M., Muttoni G., Vaillancourt B., Buell C.R., Kaeppler S.M. & de Leon N. (2013) Marker density and read depth for genotyping populations using genotyping-by-sequencing. *Genetics* **193**, 1073-81.
- Bellinge R.H.S., Liberles D.A., Iaschi S.P.A., O'Brien P.A. & Tay G.K. (2005) Myostatin and its implications on animal breeding: a review. *Animal Genetics* **36**, 1-6.
- Benedito-Palos L., Ballester-Lozano G. & Pérez-Sánchez J. (2014) Wide-gene expression analysis of lipid-relevant genes in nutritionally challenged gilthead sea bream (*Sparus aurata*). *Gene* **547**, 34-42.
- Bentsen H.B., Gjerde B., Eknath A.E., de Vera M.S.P., Velasco R.R., Danting J.C., Dionisio E.E., Longalong F.M., Reyes R.A., Abella T.A., Tayamen M.M. & Ponzoni R.W. (2017) Genetic improvement of farmed tilapias: Response to five generations of selection for increased body

- weight at harvest in *Oreochromis niloticus* and the further impact of the project. *Aquaculture* **468**, Part 1, 206-17.
- Bentsen H.B. & Olesen I. (2002) Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture* **204**, 349-59.
- Berthelot C., Brunet F., Chalopin D., Juanchich A., Bernard M., Noël B., Bento P., Da Silva C., Labadie K., Alberti A., Aury J.-M., Louis A., Dehais P., Bardou P., Montfort J., Klopp C., Cabau C., Gaspin C., Thorgaard G.H., Boussaha M., Quillet E., Guyomard R., Galiana D., Bobe J., Volff J.-N., Genêt C., Wincker P., Jaillon O., Crollius H.R. & Guiguen Y. (2014) The rainbow trout genome provides novel insights into evolution after whole-genome duplication in vertebrates. *Nature Communications* **5**, 3657.
- Bezault E., Balaesque P., Toguyeni A., Fermon Y., Araki H., Baroiller J.-F. & Rognon X. (2011) Spatial and temporal variation in population genetic structure of wild Nile tilapia (*Oreochromis niloticus*) across Africa. *BMC Genetics* **12**, 102.
- Bhatta S., Iwai T., Miura C., Higuchi M., Shimizu-Yamaguchi S., Fukada H. & Miura T. (2012) Gonads directly regulate growth in teleosts. *Proceedings of the National Academy of Sciences* **109**, 11408-12.
- Biga P.R. & Meyer J. (2009) Growth hormone differentially regulates growth and growth-related gene expression in closely related fish species. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **154**, 465-73.
- Biro P.A., Abrahams M.V., Post J.R. & Parkinson E.A. (2006) Behavioural trade-offs between growth and mortality explain evolution of submaximal growth rates. *Journal of Animal Ecology* **75**, 1165-71.
- Biscione F., Pignalberi C., Totteri A., Messina F. & Altamura G. (2007) Cardiovascular effects of Omega-3 free fatty acids. *Current Vascular Pharmacology* **5**, 163-72.
- Bizuayehu T.T. & Babiak I. (2014) MicroRNA in teleost fish. *Genome Biology and Evolution* **6**, 1911-37.
- Blanckenhorn W.U. (2005) Behavioral causes and consequences of sexual size dimorphism. *Ethology* **111**, 977-1016.
- Bligh E.G. & Dyer W.J. (1959) A rapid method for total lipid extraction and purification. *Canadian Journal of Biochemistry and Physiology* **37**, 911-7.
- Böhne A., Sengstag T. & Salzburger W. (2014) Comparative transcriptomics in east african cichlids reveals sex- and species-specific expression and new candidates for sex differentiation in fishes. *Genome Biology and Evolution* **6**, 2567-85.
- Botstein D., White R.L., Skolnick M. & Davis R.W. (1980) Construction of a genetic linkage map in man using restriction fragment length polymorphisms. *American Journal of Human Genetics* **32**, 314-31.
- Bouza C., Hermida M., Pardo B.G., Fernandez C., Fortes G.G., Castro J., Sanchez L., Presa P., Perez M. & Sanjuan A. (2007) A microsatellite genetic map of the turbot (*Scophthalmus maximus*). *Genetics* **177**.
- Bowman C.J., Ayer D.E. & Dynlacht B.D. (2014) Foxk proteins repress the initiation of starvation-induced atrophy and autophagy programs. *Nature cell biology* **16**, 1202-14.
- Bradford J.R., Hey Y., Yates T., Li Y., Pepper S.D. & Miller C.J. (2010) A comparison of massively parallel nucleotide sequencing with oligonucleotide microarrays for global transcription profiling. *BMC Genomics* **11**, 282.
- Bradley K.M., Breyer J.P., Melville D.B., Broman K.W., Knapik E.W. & Smith J.R. (2011) An SNP-based linkage map for zebrafish reveals sex determination loci. *G3: Genes, Genomes, Genetics* **1**, 3-9.
- Bradley K.M., Elmore J.B., Breyer J.P., Yaspan B.L., Jessen J.R., Knapik E.W. & Smith J.R. (2007) A major zebrafish polymorphism resource for genetic mapping. *Genome Biology* **8**, R55-R.
- Brawand D., Wagner C.E., Li Y.I., Malinsky M., Keller I., Fan S., Simakov O., Ng A.Y., Lim Z.W., Bezault E., Turner-Maier J., Johnson J., Alcazar R., Noh H.J., Russell P., Aken B., Alföldi J., Amemiya C., Azzouzi N., Baroiller J.F., Barloy-Hubler F., Berlin A., Bloomquist R., Carleton K.L., Conte M.A., D'Cotta H., Eshel O., Gaffney L., Galibert F., Gante H.F., Gnerre S., Greuter L., Guyon R., Haddad N.S., Haerty W., Harris R.M., Hofmann H.A., Hourlier T., Hulata G., Jaffe D.B., Lara M., Lee A.P., MacCallum I., Mwaiko S., Nikaido M., Nishihara H., Ozouf-Costaz C., Penman D.J., Przybylski D., Rakotomanga M., Renn S.C., Ribeiro F.J., Ron M., Salzburger W., Sanchez-Pulido L., Santos M.E., Searle S., Sharpe T., Swofford R., Tan F.J., Williams L., Young S., Yin S., Okada N., Kocher T.D., Miska E.A., Lander E.S., Venkatesh B., Fernald R.D., Meyer A., Ponting C.P., Streelman J.T., Lindblad-Toh K., Seehausen O. & Di Palma F. (2014) The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**, 375-81.
- Brenner R.R. (1981) Nutritional and hormonal factors influencing desaturation of essential fatty acids. *Progress in Lipid Research* **20**, 41-7.
- Brenner S., Johnson M., Bridgham J., Golda G., Lloyd D.H., Johnson D., Luo S., McCurdy S., Foy M., Ewan M., Roth R., George D., Eletr S., Albrecht G., Vermaas E., Williams S.R., Moon K., Burcham T., Pallas M., DuBridge R.B., Kirchner J., Fearon K., Mao J.-i. & Corcoran K. (2000) Gene expression analysis by massively parallel signature sequencing (MPSS) on microbead arrays. *Nature Biotechnology* **18**, 630-4.
- Bruford M.W. & Wayne R.K. (1993) Microsatellites and their application to population genetic studies. *Current Opinion in Genetics & Development* **3**, 939-43.

- Bush W.S. & Moore J.H. (2012) Chapter 11: Genome-Wide Association Studies. *PLOS Computational Biology* **8**, e1002822.
- Buzzi M., Henderson R.J. & Sargent J.R. (1997) Biosynthesis of docosahexaenoic acid in trout hepatocytes proceeds via 24-carbon intermediates. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **116**, 263-7.
- Cameron N.D., Enser M., Nute G.R., Whittington F.M., Penman J.C., Fiskén A.C., Perry A.M. & Wood J.D. (2000) Genotype with nutrition interaction on fatty acid composition of intramuscular fat and the relationship with flavour of pig meat. *Meat Science* **55**, 187-95.
- Campos-Ramos R., Harvey S.C., McAndrew B.J. & Penman D.J. (2003) An investigation of sex determination in the Mozambique tilapia, *Oreochromis mossambicus*, using synaptonemal complex analysis, FISH, sex reversal and gynogenesis. *Aquaculture* **221**, 125-40.
- Campos C., Valente L.M.P., Borges P., Bizuayehu T. & Fernandes J.M.O. (2010) Dietary lipid levels have a remarkable impact on the expression of growth-related genes in Senegalese sole (*Solea senegalensis* Kaup). *The Journal of Experimental Biology* **213**, 200-9.
- Canosa L.F., Chang J.P. & Peter R.E. (2007) Neuroendocrine control of growth hormone in fish. *General and Comparative Endocrinology* **151**, 1-26.
- Castaño-Sánchez C., Fuji K., Ozaki A., Hasegawa O., Sakamoto T., Morishima K., Nakayama I., Fujiwara A., Masaoka T., Okamoto H., Hayashida K., Tagami M., Kawai J., Hayashizaki Y. & Okamoto N. (2010) A second generation genetic linkage map of Japanese flounder (*Paralichthys olivaceus*). *BMC Genomics* **11**, 554.
- Castell J. (1979) Review of lipid requirements of finfish. In: *Finfish nutrition and fishfeed technology* (eds. by E.; HJ & K. T), pp. 59-84, Hamburg.
- Castillo J., Codina M., Martínez M.L., Navarro I. & Gutiérrez J. (2004) Metabolic and mitogenic effects of IGF-I and insulin on muscle cells of rainbow trout. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **286**, R935.
- Castro L.F.C., Monroig Ó., Leaver M.J., Wilson J., Cunha I. & Tocher D.R. (2012) Functional desaturase Fads1 ($\Delta 5$) and Fads2 ($\Delta 6$) orthologues evolved before the origin of jawed vertebrates. *Plos One* **7**, e31950.
- Catchen J.M., Amores A., Hohenlohe P., Cresko W. & Postlethwait J.H. (2011) Stacks: building and genotyping loci de novo from short-read sequences. *G3 (Bethesda)* **1**, 171-82.
- Cenadelli S., Maran V., Bongioni G., Fusetti L., Parma P. & Aleandri R. (2007) Identification of nuclear SNPs in gilthead seabream. *Journal of Fish Biology* **70**, 399-405.
- Chauke E., Cukrowska E., Thaela-Chimuka M.-J., Chimuka L., Nsengimana H. & Tutu H. (2008) Fatty acids composition in South African freshwater fish as indicators of food quality. *Water SA* **34**, 119-26.
- Cheng L., Liu L., Yu X., Wang D. & Tong J. (2010) A linkage map of common carp (*Cyprinus carpio*) based on AFLP and microsatellite markers. *Animal Genetics* **41**, 191-8.
- Cherfas N.B., Gomelsky B., BenDom N., Joseph D., Cohen S., Israel I., Kabessa M., Zohar G., Peretz Y., Mires D. & Hulata G. (1996) Assessment of all-female common carp progenies for fish culture. *Israeli Journal of Aquaculture-Bamidgeh* **48**, 149-57.
- Chistiakov D.A., Hellemans B., Haley C.S., Law A.S., Tsigenopoulos C.S., Kotoulas G., Bertotto D., Libertini A. & Volckaert F.A.M. (2005) A microsatellite linkage map of the european sea bass *Dicentrarchus labrax* L. *Genetics* **170**, 1821-6.
- Chistiakov D.A., Tsigenopoulos C.S., Lagnel J., Guo Y.M., Hellemans B., Haley C.S., Volckaert F.A.M. & Kotoulas G. (2008) A combined AFLP and microsatellite linkage map and pilot comparative genomic analysis of European sea bass *Dicentrarchus labrax* L. *Animal Genetics* **39**, 623-34.
- Churchill G.A. & Doerge R.W. (1994) Empirical threshold values for quantitative trait mapping. *Genetics* **138**, 963-71.
- Cleveland B.M. & Weber G.M. (2013) Effects of triploidy on growth and protein degradation in skeletal muscle during recovery from feed deprivation in juvenile rainbow trout (*Oncorhynchus mykiss*). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **166**, 128-37.
- Cnaani A., Gall G.A.E. & Hulata G. (2000) Cold tolerance of tilapia species and hybrids. *Aquaculture International* **8**, 289-98.
- Cnaani A., Hallerman E.M., Ron M., Weller J.I., Indelman M., Kashi Y., Gall G.A.E. & Hulata G. (2003) Detection of a chromosomal region with two quantitative trait loci, affecting cold tolerance and fish size, in an F2 tilapia hybrid. *Aquaculture* **223**, 117-28.
- Cnaani A., Lee B.Y., Zilberman N., Ozouf-Costaz C., Hulata G., Ron M., D'Hont A., Baroiller J.F., D'Cotta H., Penman D.J., Tomasino E., Coutanceau J.P., Pepely E., Shirak A. & Kocher T.D. (2008) Genetics of sex Determination in Tilapiine species. *Sexual Development* **2**, 43-54.
- Cnaani A., Ron M., Lee B.Y., Hulata G., Kocher T.D. & Seroussi E. (2002) Mapping the transferrin gene in tilapia. *Animal Genetics* **33**, 78-80.
- Cnaani A., Zilberman N., Tinman S., Hulata G. & Ron M. (2004) Genome-scan analysis for quantitative trait loci in an F2 tilapia hybrid. *Molecular Genetics and Genomics* **272**, 162-72.
- Codina M., García de la serrana D., Sánchez-Gurmaches J., Montserrat N., Chistyakova O., Navarro I. & Gutiérrez J. (2008) Metabolic and mitogenic effects of IGF-II in rainbow trout (*Oncorhynchus*

- mykiss*) myocytes in culture and the role of IGF-II in the PI3K/Akt and MAPK signalling pathways. *General and Comparative Endocrinology* **157**, 116-24.
- Coimbra M.R.M., Kobayashi K., Koretsugu S., Hasegawa O., Ohara E., Ozaki A., Sakamoto T., Naruse K. & Okamoto N. (2003) A genetic linkage map of the Japanese flounder, *Paralichthys olivaceus*. *Aquaculture* **220**.
- Collard B.C.Y., Jahufer M.Z.Z., Brouwer J.B. & Pang E.C.K. (2005) An introduction to markers, quantitative trait loci (QTL) mapping and marker-assisted selection for crop improvement: The basic concepts. *Euphytica* **142**, 169-96.
- Consuegra S., John E., Verspoor E. & de Leaniz C.G. (2015) Patterns of natural selection acting on the mitochondrial genome of a locally adapted fish species. *Genetics Selection Evolution* **47**, 58.
- Conte M.A., Gammerdinger W.J., Bartie K.L., Penman D.J. & Kocher T.D. (2017) A high quality assembly of the Nile Tilapia (*Oreochromis niloticus*) genome reveals the structure of two sex determination regions. *BMC Genomics* **18**, 341.
- Costa V., Angelini C., De Feis I. & Ciccodicola A. (2010) Uncovering the complexity of transcriptomes with RNA-Seq. *Journal of Biomedicine and Biotechnology* **2010**, 19.
- Danzmann R.G., Kocmarek A.L., Norman J.D., Rexroad C.E., 3rd & Palti Y. (2016) Transcriptome profiling in fast versus slow-growing rainbow trout across seasonal gradients. *BMC Genomics* **17**, 60.
- Darvasi A., Weinreb A., Minke V., Weller J.I. & Soller M. (1993) Detecting marker-QTL linkage and estimating QTL gene effect and map location using a saturated genetic map. *Genetics* **134**, 943-51.
- Davey J.W., Hohenlohe P.A., Etter P.D., Boone J.Q., Catchen J.M. & Blaxter M.L. (2011) Genome-wide genetic marker discovery and genotyping using next-generation sequencing. *Nature Review Genetics* **12**, 499-510.
- Davidson W.S., Koop B.F., Jones S.J.M., Iturra P., Vidal R., Maass A., Jonassen I., Lien S. & Omholt S.W. (2010) Sequencing the genome of the Atlantic salmon (*Salmo salar*). *Genome Biol* **11**.
- Davis D.A. (2015) *Feed and Feeding Practices in Aquaculture*. Woodhead Publishing, Oxford.
- Davis L.K., Pierce A.L., Hiramatsu N., Sullivan C.V., Hirano T. & Grau E.G. (2008) 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**, 544-51.
- De-Santis C. & Jerry D.R. (2007) Candidate growth genes in finfish — Where should we be looking? *Aquaculture* **272**, 22-38.
- Degani G., Tzchori I., Yom-Din S., Goldberg D. & Jackson K. (2003) Growth differences and growth hormone expression in male and female European eels [*Anguilla anguilla* (L.)]. *General and Comparative Endocrinology* **134**, 88-93.
- Derayat A., Houston R.D., Guy D.R., Hamilton A., Ralph J., Spreckley N., Taggart J.B., McAndrew B.J. & Haley C.S. (2007) Mapping QTL affecting body lipid percentage in Atlantic salmon (*Salmo salar*). *Aquaculture* **272**, **Supplement 1**, S250-S1.
- Devlin R.H., Biagi C.A., Yesaki T.Y., Smailus D.E. & Byatt J.C. (2001) Growth of domesticated transgenic fish. *Nature* **409**, 781-2.
- Devlin R.H., Sakhrani D., White S. & Overturf K. (2013) Effects of domestication and growth hormone transgenesis on mRNA profiles in rainbow trout (*Oncorhynchus mykiss*). *Journal of Animal Science* **91**, 5247-58.
- Doerge R.W. (2002) Mapping and analysis of quantitative trait loci in experimental populations. *Nature Review Genetics* **3**, 43-52.
- Doerge R.W. & Rebai A. (1996) Significance thresholds for QTL interval mapping tests. *Heredity* **76**, 459-64.
- Doudet T. (1992) Brackishwater tolerance of some species and hybrids of *Oreochromis* for use in lagoon aquaculture (Ivory Coast). *Aquaculture* **102**, 275-88.
- Du Z.Y., Liu Y.J., Tian L.X., Wang J.T., Wang Y. & Liang G.Y. (2005) Effect of dietary lipid level on growth, feed utilization and body composition by juvenile grass carp (*Ctenopharyngodon idella*). *Aquaculture Nutrition* **11**, 139-46.
- Duan C., Ren H. & Gao S. (2010) Insulin-like growth factors (IGFs), IGF receptors, and IGF-binding proteins: roles in skeletal muscle growth and differentiation. *General and Comparative Endocrinology* **167**, 344-51.
- El-Sayed A.F.M. (2006) *Tilapia culture*. CABI Publishing, London, Oxford.
- El-Sayed A.F.M., Mansour C.R. & Ezzat A.A. (2005) Effects of dietary lipid source on spawning performance of Nile tilapia (*Oreochromis niloticus*) broodstock reared at different water salinities. *Aquaculture* **248**, 187-96.
- Elshire R.J., Glaubitz J.C., Sun Q., Poland J.A., Kawamoto K., Buckler E.S. & Mitchell S.E. (2011) A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *Plos One* **6**, e19379.
- Embry G.C. & Hayford C.O. (1925) The advantage of rearing brook trout fingerlings from selected breeders. *Transactions of the American Fisheries Society* **55**, 135-48.
- Eshel O., Shirak A., Dor L., Band M., Zak T., Markovich-Gordon M., Chalifa-Caspi V., Feldmesser E., Weller J.I., Seroussi E., Hulata G. & Ron M. (2014) Identification of male-specific amh

- duplication, sexually differentially expressed genes and microRNAs at early embryonic development of Nile tilapia (*Oreochromis niloticus*). *BMC Genomics* **15**, 774.
- Eshel O., Shirak A., Weller J.I., Slossman T., Hulata G., Cnaani A. & Ron M. (2011) Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Animal Genetics* **42**, 222-4.
- Evans D.M. & Cardon L.R. (2004) Guidelines for genotyping in genome-wide linkage studies: Single-nucleotide polymorphism maps versus microsatellite maps. *The American Journal of Human Genetics* **75**, 687-92.
- Ezaz M.T., Harvey S.C., Boonphakdee C., Teale A.J., McAndrew B.J. & Penman D.J. (2004) Isolation and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus* L.). *Marine Biotechnology* **6**, 435-45.
- Fairbairn D.J. (1997) Allometry for sexual size dimorphism: Pattern and process in the coevolution of body size in males and females. *Annual Review of Ecology and Systematics* **28**, 659-87.
- Falconer D.S. & Mackay T.F. (1996) Introduction to quantitative genetics (4th edn). In: *Trends in Genetics*. Benjamin Cummings, Harlow, United Kingdom.
- FAO (2014) World aquaculture production by species groups. Fisheries and Aquaculture Department.
- FAO (2016) The state of world fisheries and aquaculture. Rome.
- FAOSTAT (2017) FAOSTAT aquaculture production
- Fauconneau B., Alami-Durante H., Laroche M., Marcel J. & Vallot D. (1995) Growth and meat quality relations in carp. *Aquaculture* **129**, 265-97.
- Fernandes J.M., MacKenzie M.G., Kinghorn J.R. & Johnston I.A. (2007) FoxK1 splice variants show developmental stage-specific plasticity of expression with temperature in the tiger pufferfish. *Journal of experimental biology* **210**, 3461-72.
- Figueras A., Robledo D., Corvelo A., Hermida M., Pereiro P., Rubiolo J.A., Gómez-Garrido J., Carreté L., Bello X., Gut M., Gut I.G., Marcet-Houben M., Forn-Cuní G., Galán B., García J.L., Abal-Fabeiro J.L., Pardo B.G., Taboada X., Fernández C., Vlasova A., Hermoso-Pulido A., Guigó R., Álvarez-Díos J.A., Gómez-Tato A., Viñas A., Maside X., Gabaldón T., Novoa B., Bouza C., Alioto T. & Martínez P. (2016) Whole genome sequencing of Turbot (*Scophthalmus maximus*; *Pleuronectiformes*): A fish adapted to demersal life. *DNA Research* **23**, 181-92.
- Franch R., Louro B., Tsalavouta M., Chatziplis D., Tsigenopoulos C.S., Sarropoulou E., Antonello J., Magoulas A., Mylonas C.C., Babbucci M., Patarnello T., Power D.M., Kotoulas G. & Bargelloni L. (2006) A genetic linkage map of the hermaphrodite teleost fish *Sparus aurata* L. *Genetics* **174**, 851-61.
- Froese R. (2006) Cube law, condition factor and weight-length relationships: history, meta-analysis and recommendations. *Journal of Applied Ichthyology* **22**, 241-53.
- Frommen J.G. & Bakker T.C.M. (2006) Inbreeding avoidance through non-random mating in sticklebacks. *Biology Letters* **2**, 232-5.
- Frøyland L., Madsen L., Eckhoff K.M., Lie Ø. & Berge R.K. (1998) Carnitine palmitoyltransferase I, carnitine palmitoyltransferase II, and Acyl-CoA oxidase activities in Atlantic salmon (*Salmo salar*). *Lipids* **33**, 923-30.
- Gao B., Radaeva S. & Park O. (2009) Liver natural killer and natural killer T cells: immunobiology and emerging roles in liver diseases. *Journal of Leukocyte Biology* **86**, 513-28.
- Geldermann H. (1975) Investigations on inheritance of quantitative characters in animals by gene markers I. Methods. *Theoretical and Applied Genetics* **46**, 319-30.
- Geng X., Liu S., Yao J., Bao L., Zhang J., Li C., Wang R., Sha J., Zeng P., Zhi D. & Liu Z. (2016) A genome-wide association study identifies multiple regions associated with head size in catfish. *G3: Genes/Genomes/Genetics* **6**, 3389-98.
- Geng X., Sha J., Liu S., Bao L., Zhang J., Wang R., Yao J., Li C., Feng J., Sun F., Sun L., Jiang C., Zhang Y., Chen A., Dunham R., Zhi D. & Liu Z. (2015) A genome-wide association study in catfish reveals the presence of functional hubs of related genes within QTLs for columnaris disease resistance. *BMC Genomics* **16**, 196.
- Gharbi K., Gautier A., Danzmann R.G., Gharbi S., Sakamoto T., Høyheim B., Taggart J.B., Cairney M., Powell R., Krieg F., Okamoto N., Ferguson M.M., Holm L.-E. & Guyomard R. (2006) A linkage map for brown trout (*Salmo trutta*): Chromosome homeologies and comparative genome organization with other salmonid fish. *Genetics* **172**, 2405-19.
- Gilbey J., Verspoor E., McLay A. & Houlihan D. (2004) A microsatellite linkage map for Atlantic salmon (*Salmo salar*). *Animal Genetics* **35**.
- Ginot F., Bordelais I., Nguyen S. & Gyapay G. (1996) Correction of some genotyping errors in automated fluorescent microsatellite analysis by enzymatic removal of one base overhangs. *Nucleic Acids Research* **24**, 540-1.
- Gjedrem T. (1997) Flesh quality improvement in fish through breeding. *Aquaculture International* **5**, 197-206.
- Gjedrem T. (2010) The first family-based breeding program in aquaculture. *Reviews in Aquaculture* **2**, 2-15.
- Gjedrem T. (2012) Genetic improvement for the development of efficient global aquaculture: A personal opinion review. *Aquaculture* **344-349**, 12-22.

- Gjedrem T. & Baranski M. (2009) Selective breeding in aquaculture: An introduction. In: *Reviews: Methods and Technologies in Fish Biology and Fisheries*. Springer Netherlands.
- Gjedrem T., Robinson N. & Rye M. (2012) The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture* **350-353**, 117-29.
- Gjerde B. & Schaeffer L.R. (1989) Body traits in rainbow trout. *Aquaculture* **80**, 25-44.
- Glasauer S.M.K. & Neuhauss S.C.F. (2014) Whole-genome duplication in teleost fishes and its evolutionary consequences. *Molecular Genetics and Genomics* **289**, 1045-60.
- Glencross B.D. (2009) Exploring the nutritional demand for essential fatty acids by aquaculture species. *Reviews in Aquaculture* **1**, 71-124.
- Goddard M.E. & Hayes B.J. (2009) Mapping genes for complex traits in domestic animals and their use in breeding programmes. *Nature Reviews Genetics* **10**, 381-91.
- Gonzalez R.J. (2012) The physiology of hyper-salinity tolerance in teleost fish: a review. *Journal of Comparative Physiology B* **182**, 321-9.
- Götz S., García-Gómez J.M., Terol J., Williams T.D., Nagaraj S.H., Nueda M.J., Robles M., Talón M., Dopazo J. & Conesa A. (2008) High-throughput functional annotation and data mining with the Blast2GO suite. *Nucleic Acids Research* **36**, 3420-35.
- Goudie C.A., Simco B.A., Davis K.B. & Carmichael G.J. (1994) Growth of channel catfish in mixed sex and monosex pond culture. *Aquaculture* **128**, 97-104.
- Goymer P. (2007) Synonymous mutations break their silence. *Nat Rev Genet* **8**, 92-.
- Grabherr M.G., Haas B.J., Yassour M., Levin J.Z., Thompson D.A., Amit I., Adiconis X., Fan L., Raychowdhury R., Zeng Q., Chen Z., Mauceli E., Hacohen N., Gnirke A., Rhind N., di Palma F., Birren B.W., Nusbaum C., Lindblad-Toh K., Friedman N. & Regev A. (2011) Trinity: reconstructing a full-length transcriptome without a genome from RNA-Seq data. *Nature Biotechnology* **29**, 644-52.
- Gregory T.R. (2001) Animal genome size database. TR Gregory.
- Grigorakis K., Alexis M.N., Taylor K.D.A. & Hole M. (2002) Comparison of wild and cultured gilthead sea bream (*Sparus aurata*); composition, appearance and seasonal variations. *International Journal of Food Science and Technology* **37**, 477-84.
- Gross R. & Nilsson J. (1999) Restriction fragment length polymorphism at the growth hormone 1 gene in Atlantic salmon (*Salmo salar* L.) and its association with weight among the offspring of a hatchery stock. *Aquaculture* **173**, 73-80.
- Gupta M.V. & Acosta B.O. (2004) A review of global tilapia farming practices. *Aquaculture Asia* **9**, 7-12.
- Gutierrez A.P., Lubieniecki K.P., Davidson E.A., Lien S., Kent M.P., Fukui S., Withler R.E., Swift B. & Davidson W.S. (2012) Genetic mapping of quantitative trait loci (QTL) for body-weight in Atlantic salmon (*Salmo salar*) using a 6.5 K SNP array. *Aquaculture* **358-359**, 61-70.
- Gutierrez A.P., Yanez J.M., Fukui S., Swift B. & Davidson W.S. (2015) Genome-wide association study (GWAS) for growth rate and age at sexual maturation in Atlantic salmon (*Salmo salar*). *Plos One* **10**, e0119730.
- Guyon R., Rakotomanga M., Azzouzi N., Coutanceau J.P., Bonillo C., D'Cotta H., Pepey E., Soler L., Rodier-Goud M., D'Hont A., Conte M.A., van Bers N.E., Penman D.J., Hitte C., Crooijmans R.P., Kocher T.D., Ozouf-Costaz C., Baroiller J.F. & Galibert F. (2012) A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* **13**, 222.
- Haas B.J., Papanicolaou A., Yassour M., Grabherr M., Blood P.D., Bowden J., Couger M.B., Eccles D., Li B., Lieber M., MacManes M.D., Ott M., Orvis J., Pochet N., Strozzi F., Weeks N., Westerman R., William T., Dewey C.N., Henschel R., LeDuc R.D., Friedman N. & Regev A. (2013) De novo transcript sequence reconstruction from RNA-seq using the Trinity platform for reference generation and analysis. *Nature Protocols* **8**, 1494-512.
- Hackett C.A. & Broadfoot L.B. (2003) Effects of genotyping errors, missing values and segregation distortion in molecular marker data on the construction of linkage maps. *Heredity* **90**, 33-8.
- Hagemeister A.L. & Sheridan M.A. (2008) Somatostatin inhibits hepatic growth hormone receptor and insulin-like growth factor I mRNA expression by activating the ERK and PI3K signaling pathways. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **295**, R490.
- Hagen Ø., Solberg C. & Johnston I.A. (2006) Sexual dimorphism of fast muscle fibre recruitment in farmed Atlantic halibut (*Hippoglossus hippoglossus* L.). *Aquaculture* **261**, 1222-9.
- Hahn M., Wilhelm J. & Pingoud A. (2001) Influence of fluorophor dye labels on the migration behavior of polymerase chain reaction--amplified short tandem repeats during denaturing capillary electrophoresis. *Electrophoresis* **22**, 2691-700.
- Haidle L., Janssen J.E., Gharbi K., Moghadam H.K., Ferguson M.M. & Danzmann R.G. (2008) Determination of quantitative trait loci (QTL) for early maturation in rainbow trout (*Oncorhynchus mykiss*). *Marine Biotechnology* **10**, 579-92.
- Haley C. & De Koning D.J. (2006) Genetical genomics in livestock: potentials and pitfalls. *Animal Genetics* **37**, 10-2.
- Haley C.S. & Knott S.A. (1992) A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity* **69**, 315-24.

- Hamady M., Walker J.J., Harris J.K., Gold N.J. & Knight R. (2008) Error-correcting barcoded primers for pyrosequencing hundreds of samples in multiplex. *Nature Methods* **5**, 235-7.
- Hamzah A., Nguyen N.H., Mekki W., Ponzoni R.W., Khaw H.L., Yee H.Y., Abu Bakar K.R. & Mohd Nor S.A. (2016) Flesh characteristics: estimation of genetic parameters and correlated responses to selection for growth rate in the GIFT strain. *Aquaculture Research* **47**, 2139-49.
- Han C., Wen X., Zheng Q. & Li H. (2011) Effect of starvation on activities and mRNA expression of lipoprotein lipase and hormone-sensitive lipase in tilapia (*Oreochromis niloticus* × *O. aureus*). *Fish Physiology and Biochemistry* **37**, 113-22.
- Hardie D.G. (2004) The AMP-activated protein kinase pathway – new players upstream and downstream. *Journal of Cell Science* **117**, 5479-87.
- Harker N., Rampling L.R., Shariflou M.R., Hayden M.J., Holton T.A., Morell M.K., Sharp P.J., Henry R.J. & Edwards K.J. (2001) Microsatellites as markers for Australian wheat improvement. *Australian Journal of Agricultural Research* **52**, 1121-30.
- Harvey K. & Tapon N. (2007) The salvador-warts-hippo pathway: An emerging tumour-suppressor network. *Nature Reviews Cancer* **7**, 182-91.
- Hastings N., Agaba M., Tocher D.R., Leaver M.J., Dick J.R., Sargent J.R. & Teale A.J. (2001) A vertebrate fatty acid desaturase with $\Delta 5$ and $\Delta 6$ activities. *Proceedings of the National Academy of Sciences* **98**, 14304-9.
- Hastings N., Agaba M.K., Tocher D.R., Zheng X., Dickson C.A., Dick J.R. & Teale A.J. (2004) Molecular cloning and functional characterization of fatty acyl desaturase and elongase cDNAs involved in the production of eicosapentaenoic and docosahexaenoic acids from α -linolenic acid in Atlantic Salmon (*Salmo salar*). *Marine Biotechnology* **6**, 463-74.
- Hayes B., Laerdahl J.K., Lien S., Moen T., Berg P., Hindar K., Davidson W.S., Koop B.F., Adzhubei A. & Høyheim B. (2007) An extensive resource of single nucleotide polymorphism markers associated with Atlantic salmon (*Salmo salar*) expressed sequences. *Aquaculture* **265**, 82-90.
- Haynes W. (2013) Bonferroni correction. In: *Encyclopedia of Systems Biology* (eds. by Dubitzky W, Wolkenhauer O, Cho K-H & Yokota H), pp. 154-. Springer New York, New York, NY.
- He A.-Y., Ning L.-J., Chen L.-Q., Chen Y.-L., Xing Q., Li J.-M., Qiao F., Li D.-L., Zhang M.-L. & Du Z.-Y. (2015) Systemic adaptation of lipid metabolism in response to low- and high-fat diet in Nile tilapia (*Oreochromis niloticus*). *Physiological Reports* **3**, e12485.
- He A., Luo Y., Yang H., Liu L., Li S. & Wang C. (2011) Complete mitochondrial DNA sequences of the Nile tilapia (*Oreochromis niloticus*) and Blue tilapia (*Oreochromis aureus*): genome characterization and phylogeny applications. *Molecular Biology Reports* **38**, 2015-21.
- Hearn T.L., Sgoutas S.A., Hearn J.A. & Sgoutas D.S. (1987) Polyunsaturated fatty acids and fat in fish flesh for selecting species for health benefits. *Journal of Food Science* **52**, 1209-11.
- Hedrick A.V. & Temeles E.J. (1989) The evolution of sexual dimorphism in animals: Hypotheses and tests. *Trends in Ecology & Evolution* **4**, 136-8.
- Heffelfinger C., Fragoso C.A., Moreno M.A., Overton J.D., Mottinger J.P., Zhao H., Tohme J. & Dellaporta S.L. (2014) Flexible and scalable genotyping-by-sequencing strategies for population studies. *BMC Genomics* **15**, 979.
- Henderson R.J. & Tocher D.R. (1987) The lipid composition and biochemistry of freshwater fish. *Progress in Lipid Research* **26**, 281-347.
- Henning F., Lee H.J., Franchini P. & Meyer A. (2014) Genetic mapping of horizontal stripes in Lake Victoria cichlid fishes: benefits and pitfalls of using RAD markers for dense linkage mapping. *Molecular Ecology* **23**, 5224-40.
- Henry C.A. & Amacher S.L. (2004) Zebrafish slow muscle cell migration induces a wave of fast muscle morphogenesis. *Developmental Cell* **7**, 917-23.
- Herpin A. & Schartl M. (2011) Dmrt1 genes at the crossroads: a widespread and central class of sexual development factors in fish. *FEBS Journal* **278**, 1010-9.
- Hinits Y. & Hughes S.M. (2007) Mef2s are required for thick filament formation in nascent muscle fibres. *Development* **134**, 2511.
- Hinits Y., Osborn D.P.S. & Hughes S.M. (2009) Differential requirements for myogenic regulatory factors distinguish medial and lateral somitic, cranial and fin muscle fibre populations. *Development* **136**, 403.
- Hjalti-Jákupsstovu S. & Haug T. (1988) Growth, sexual maturation, and spawning season of Atlantic halibut, *Hippoglossus hippoglossus*, in Faroese waters. *Fisheries Research* **6**, 201-15.
- Hohenlohe P.A., Day M.D., Amish S.J., Miller M.R., Kamps-Hughes N., Boyer M.C., Muhlfeld C.C., Allendorf F.W., Johnson E.A. & Luikart G. (2013) Genomic patterns of introgression in rainbow and westslope cutthroat trout illuminated by overlapping paired-end RAD sequencing. *Molecular Ecology* **22**, 3002-13.
- Holman R.T. (1998) The slow discovery of the importance of $\omega 3$ essential fatty acids in human health. *The Journal of Nutrition* **128**, 427S-33S.
- Holub D.J. & Holub B.J. (2004) Omega-3 fatty acids from fish oils and cardiovascular disease. *Molecular and Cellular Biochemistry* **263**, 217-25.
- Hong H., Xu L., Liu J., Jones W.D., Su Z., Ning B., Perkins R., Ge W., Miclaus K., Zhang L., Park K., Green B., Han T., Fang H., Lambert C.G., Vega S.C., Lin S.M., Jafari N., Czika W., Wolfinger

- R.D., Goodsaid F., Tong W. & Shi L. (2012) Technical reproducibility of genotyping SNP arrays used in genome-wide association studies. *Plos One* **7**, e44483.
- Houston R.D., Taggart J.B., Cézard T., Bekaert M., Lowe N.R., Downing A., Talbot R., Bishop S.C., Archibald A.L., Bron J.E., Penman D.J., Davassi A., Brew F., Tinch A.E., Gharbi K. & Hamilton A. (2014) Development and validation of a high density SNP genotyping array for Atlantic salmon (*Salmo salar*). *BMC Genomics* **15**, 90.
- Hu G., Gu W., Sun P., Bai Q. & Wang B. (2016a) Transcriptome analyses reveal lipid metabolic process in liver related to the difference of carcass fat content in Rainbow trout (*Oncorhynchus mykiss*). *International Journal of Genomics* **2016**, 10.
- Hu P., Liu M., Liu Y., Wang J., Zhang D., Niu H., Jiang S., Wang J., Zhang D., Han B., Xu Q. & Chen L. (2016b) Transcriptome comparison reveals a genetic network regulating the lower temperature limit in fish. **6**, 28952.
- Huang C.W., Li Y.H., Hu S.Y., Chi J.R., Lin G.H., Lin C.C., Gong H.Y., Chen J.Y., Chen R.H., Chang S.J., Liu F.G. & Wu J.L. (2012) Differential expression patterns of growth-related microRNAs in the skeletal muscle of Nile tilapia (*Oreochromis niloticus*). *Journal of Animal Science* **90**, 4266-79.
- Huang X. & Madan A. (1999) CAP3: A DNA sequence assembly program. *Genome Research* **9**, 868-77.
- Hubert S. & Hedgecock D. (2004) Linkage maps of microsatellite DNA markers for the pacific oyster *Crassostrea gigas*. *Genetics* **168**.
- Hubert S., Higgins B., Borza T. & Bowman S. (2010) Development of a SNP resource and a genetic linkage map for Atlantic cod (*Gadus morhua*). *BMC Genomics* **11**, 191.
- Hulata G., Wohlfarth G.W. & Halevy A. (1986) Mass selection for growth rate in the Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **57**, 177-84.
- İbrahim Haliloğlu H., Bayır A., Necdet Sirkecioğlu A., Mevlüt Aras N. & Atamanalp M. (2004) Comparison of fatty acid composition in some tissues of rainbow trout (*Oncorhynchus mykiss*) living in seawater and freshwater. *Food Chemistry* **86**, 55-9.
- Inoue J., Sato Y., Sinclair R., Tsukamoto K. & Nishida M. (2015) Rapid genome reshaping by multiple-gene loss after whole-genome duplication in teleost fish suggested by mathematical modeling. *Proceedings of the National Academy of Sciences of the United States of America* **112**, 14918-23.
- Jackson R.J., Hellen C.U.T. & Pestova T.V. (2010) The mechanism of eukaryotic translation initiation and principles of its regulation. *Nature Reviews Molecular Cell Biology* **11**, 113.
- Jansen R.C. (1993) Interval mapping of multiple quantitative trait loci. *Genetics* **135**, 205-11.
- Jansen R.C. & Stam P. (1994) High resolution of quantitative traits into multiple loci via interval mapping. *Genetics* **136**, 1447-55.
- Jauncey K. (2000) Nutritional requirements. In: *Tilapias: biology and exploitation* (pp. 327-75. Springer.
- Ji X.-S., Liu H.-W., Chen S.-L., Jiang Y.-L. & Tian Y.-S. (2011) Growth differences and dimorphic expression of growth hormone (GH) in female and male *Cynoglossus semilaevis* after male sexual maturation. *Marine Genomics* **4**, 9-16.
- Jiang J., Feng L., Tang L., Liu Y., Jiang W. & Zhou X. (2015) Growth rate, body composition, digestive enzymes and transaminase activities, and plasma ammonia concentration of different weight Jian carp (*Cyprinus carpio* var. *Jian*). *Animal Nutrition* **1**, 373-7.
- Jiao B., Huang X., Chan C.B., Zhang L., Wang D. & Cheng C.H. (2006) The co-existence of two growth hormone receptors in teleost fish and their differential signal transduction, tissue distribution and hormonal regulation of expression in seabream. *Journal of Molecular Endocrinology* **36**, 23-40.
- Jobling M. (2007) Nutrient partitioning and the influence of feed composition on body composition. In: *Food Intake in Fish* (pp. 354-76. Blackwell Science Ltd.
- Johnston I.A. (1999) Muscle development and growth: potential implications for flesh quality in fish. *Aquaculture* **177**, 99-115.
- Kacperczyk A., Jagla T. & Daczewska M. (2009) Pax-3 and Pax-7 Label Muscle Progenitor Cells During Myotomal Myogenesis in *Coregonus lavaretus* (Teleostei: Coregonidae). *Anatomia, Histologia, Embryologia* **38**, 411-8.
- Kacperczyk A., Jędrzejowska I. & Daczewska M. (2011) Differentiation and growth of myotomal muscles in a non-model tropical fish *Pterophyllum scalare* (Teleostei: Cichlidae). *Anatomia, Histologia, Embryologia* **40**, 411-8.
- Kamal A.H.M.M. & Mair G.C. (2005) Salinity tolerance in superior genotypes of tilapia, *Oreochromis niloticus*, *Oreochromis mossambicus* and their hybrids. *Aquaculture* **247**, 189-201.
- Kamei H., Ding Y., Kajimura S., Wells M., Chiang P. & Duan C. (2011) Role of IGF signaling in catch-up growth and accelerated temporal development in zebrafish embryos in response to oxygen availability. *Development* **138**, 777.
- Karamichou E., Richardson R.I., Nute G.R., Gibson K.P. & Bishop S.C. (2006) Genetic analyses and quantitative trait loci detection, using a partial genome scan, for intramuscular fatty acid composition in Scottish Blackface sheep. *Journal of Animal Science* **84**, 3228-38.
- Karapanagiotidis I.T. (2017) Nutrient profiles of tilapia. In: *Tilapia in Intensive Co-culture* (pp. 261-305. John Wiley & Sons, Ltd.
- Karapanagiotidis I.T., Bell M.V., Little D.C. & Yakupitiyage A. (2007) Replacement of dietary fish oils by alpha-linolenic acid-rich oils lowers omega 3 content in Tilapia flesh. *Lipids* **42**, 547-59.

- Karapanagiotidis I.T., Bell M.V., Little D.C., Yakupitiyage A. & Rakshit S.K. (2006) Polyunsaturated fatty acid content of wild and farmed tilapias in thailand: Effect of aquaculture practices and implications for human nutrition. *Journal of Agricultural and Food Chemistry* **54**, 4304-10.
- Katagiri T., Kidd C., Tomasino E., Davis J.T., Wishon C., Stern J.E., Carleton K.L., Howe A.E. & Kocher T.D. (2005) A BAC-based physical map of the Nile tilapia genome. *BMC Genomics* **6**, 89.
- Kause A., Quinton C., Airaksinen S., Ruohonen K. & Koskela J. (2011) Quality and production trait genetics of farmed European whitefish, *Coregonus lavaretus*. *Journal of Animal Science* **89**, 959-71.
- Kause A., Ritola O., Paananen T., Mäntysaari E. & Eskelinen U. (2002) Coupling body weight and its composition: a quantitative genetic analysis in rainbow trout. *Aquaculture* **211**, 65-79.
- Kaushik S.J. (1998) Nutritional bioenergetics and estimation of waste production in non-salmonids. *Aquat. Living Resour.* **11**, 211-7.
- Kearsey M.J. & Pooni H.S. (1998) *The genetical analysis of quantitative traits*. Stanley Thornes (Publishers) Ltd.
- Keller I., Wagner C.E., Greuter L., Mwaiko S., Selz O.M., Sivasundar A., Wittwer S. & Seehausen O. (2013) Population genomic signatures of divergent adaptation, gene flow and hybrid speciation in the rapid radiation of Lake Victoria cichlid fishes. *Molecular Ecology* **22**, 2848-63.
- Kiessling A., Pickova J., Johansson L., Asgard T., Storebakken T. & Kiessling K.H. (2001) Changes in fatty acid composition in muscle and adipose tissue of farmed rainbow trout (*Oncorhynchus mykiss*) in relation to ration and age. *Food Chemistry* **73**, 271-84.
- Kinaro Z.O., Xue L., Nyaundi K.J. & Shen J. (2015) The mitochondrial genome of an endangered native Singidia tilapia, *Oreochromis esculentus*: genome organization and control region polymorphism. *Mitochondrial DNA Part A* **27**, 4344-6.
- Kincaid H.L. (1983) Inbreeding in fish populations used for aquaculture. *Aquaculture* **33**, 215-27.
- Klett V. & Meyer A. (2002) What, if anything, is a Tilapia?—Mitochondrial ND2 phylogeny of tilapiines and the evolution of parental care systems in the African cichlid fishes. *Molecular Biology and Evolution* **19**, 865-83.
- Kobayashi M., Msangi S., Batka M., Vannuccini S., Dey M.M. & Anderson J.L. (2015) Fish to 2030: The role and opportunity for aquaculture. *Aquaculture Economics & Management* **19**, 282-300.
- Kobelkowsky A. (2004) Sexual dimorphism of the flounder *Bothus robinsi* (Pisces: Bothidae). *Journal of Morphology* **260**, 165-71.
- Kocher T.D., Lee W.J., Sobolewska H., Penman D. & McAndrew B. (1998) A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* **148**, 1225-32.
- Kocmarek A.L., Ferguson M.M. & Danzmann R.G. (2015) Co-localization of growth QTL with differentially expressed candidate genes in rainbow trout. *Genome* **58**, 393-403.
- Kocour M., Mauger S., Rodina M., Gela D., Linhart O. & Vandeputte M. (2007) Heritability estimates for processing and quality traits in common carp (*Cyprinus carpio* L.) using a molecular pedigree. *Aquaculture* **270**, 43-50.
- Kolditz C.-I., Paboeuf G., Borthaire M., Esquerré D., SanCristobal M., Lefèvre F. & Médale F. (2008) Changes induced by dietary energy intake and divergent selection for muscle fat content in rainbow trout (*Oncorhynchus mykiss*), assessed by transcriptome and proteome analysis of the liver. *BMC Genomics* **9**, 506-.
- Kolditz C., Lefèvre F., Borthaire M. & Médale F. (2007) Transcriptome and proteome analysis of changes induced in trout liver by suppression of dietary fish oil. *The FASEB Journal* **21**, A1402-A3.
- Kris-Etherton P.M., Harris W.S. & Appel L.J. (2002) Fish consumption, fish oil, Omega-3 fatty acids, and cardiovascular disease. *Circulation* **106**, 2747.
- Kuang Y., Zheng X., Lv W., Cao D. & Sun X. (2015) Mapping quantitative trait loci for flesh fat content in common carp (*Cyprinus carpio*). *Aquaculture* **435**, 100-5.
- Kukurba K.R. & Montgomery S.B. (2015) RNA sequencing and analysis. *Cold Spring Harbor protocols* **2015**, 951-69.
- Küttner E., Moghadam H.K., Skúlason S., Danzmann R.G. & Ferguson M.M. (2011) Genetic architecture of body weight, condition factor and age of sexual maturation in Icelandic Arctic charr (*Salvelinus alpinus*). *Molecular Genetics and Genomics* **286**, 67-79.
- LaFramboise T. (2009) Single nucleotide polymorphism arrays: a decade of biological, computational and technological advances. *Nucleic Acids Research* **37**, 4181-93.
- Laghari M.Y., Lashari P., Zhang X., Xu P., Narejo N.T., Xin B., Zhang Y. & Sun X. (2015) QTL mapping for economically important traits of common carp (*Cyprinus carpio* L.). *Journal of Applied Genetics* **56**, 65-75.
- Lallias D., Beaumont A.R., Haley C.S., Boudry P., Heurtebise S. & Lapègue S. (2007) A first-generation genetic linkage map of the european flat oyster *Ostrea edulis* (L.) based on AFLP and microsatellite markers. *Animal Genetics* **38**, 560-8.
- Lam S.H., Lui E.Y., Li Z., Cai S., Sung W.-K., Mathavan S., Lam T.J. & Ip Y.K. (2014) Differential transcriptomic analyses revealed genes and signaling pathways involved in iono-osmoregulation and cellular remodeling in the gills of euryhaline Mozambique tilapia, *Oreochromis mossambicus*. *BMC Genomics* **15**, 921.
- Lande R. & Thompson R. (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* **124**, 743-56.

- Lander E. & Kruglyak L. (1995) Genetic dissection of complex traits: guidelines for interpreting and reporting linkage results. *Nature Genetics* **11**, 241-7.
- Lander E.S. & Botstein D. (1989) Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics* **121**, 185-99.
- Laplante M. & Sabatini D.M. (2009) mTOR signaling at a glance. *Journal of Cell Science* **122**, 3589-94.
- Leaf A. & Weber P.C. (1988) Cardiovascular effects of n-3 fatty acids. *New England Journal of Medicine* **318**, 549-57.
- Leaver M.J., Taggart J.B., Villeneuve L., Bron J.E., Guy D.R., Bishop S.C., Houston R.D., Matika O. & Tocher D.R. (2011) Heritability and mechanisms of n-3 long chain polyunsaturated fatty acid deposition in the flesh of Atlantic salmon. *Comparative Biochemistry and Physiology Part D: Genomics and Proteomics* **6**, 62-9.
- Leclercq E., Taylor J.F., Hunter D. & Migaud H. (2010) Body size dimorphism of sea-reared Atlantic salmon (*Salmo salar* L.): Implications for the management of sexual maturation and harvest quality. *Aquaculture* **301**, 47-56.
- Lee B.-Y., Howe A.E., Conte M.A., D'Cotta H., Pepey E., Baroiller J.-F., di Palma F., Carleton K.L. & Kocher T.D. (2010) An EST resource for tilapia based on 17 normalized libraries and assembly of 116,899 sequence tags. *BMC Genomics* **11**, 278.
- Lee B.-Y., Coutanceau J.P., Ozouf-Costaz C., D'Cotta H., Baroiller J.F. & Kocher T.D. (2011) Genetic and physical mapping of sex-linked AFLP markers in Nile tilapia (*Oreochromis niloticus*). *Marine Biotechnology* **13**, 557-62.
- Lee B.-Y., Hulata G. & Kocher T.D. (2004) Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* **92**, 543-9.
- Lee B.-Y., Lee W.J., Streelman J.T., Carleton K.L., Howe A.E., Hulata G., Slettan A., Stern J.E., Terai Y. & Kocher T.D. (2005) A second-generation genetic linkage map of tilapia (*Oreochromis spp.*). *Genetics* **170**, 237-44.
- Lee B.-Y., Penman D.J. & Kocher T.D. (2003) Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal Genetics* **34**, 379-83.
- Lee S.-J., Huynh T.V., Lee Y.-S., Sebald S.M., Wilcox-Adelman S.A., Iwamori N., Lepper C., Matzuk M.M. & Fan C.-M. (2012) Role of satellite cells versus myofibers in muscle hypertrophy induced by inhibition of the myostatin/activin signaling pathway. *Proceedings of the National Academy of Sciences* **109**, E2353-E60.
- Lefevre F., Cardinal M., Bugeon J., Labbe L., Medale F. & Quillet E. (2015) Selection for muscle fat content and triploidy affect flesh quality in pan-size rainbow trout, *Oncorhynchus mykiss*. *Aquaculture* **448**, 569-77.
- Lemarié G., Baroiller J.F., Clota F., Lazard J. & Dosdat A. (2004) A simple test to estimate the salinity resistance of fish with specific application to *O. niloticus* and *S. melanotheron*. *Aquaculture* **240**, 575-87.
- Lenormand T. & Dutheil J. (2005) Recombination difference between sexes: A role for haploid selection. *PLOS Biology* **3**, e63.
- Li G., Sinclair A.J. & Li D. (2011) Comparison of lipid content and fatty acid composition in the edible meat of wild and cultured freshwater and marine fish and shrimps from China. *Journal of Agricultural and Food Chemistry* **59**, 1871-81.
- Li H.-O. & Yamada J. (1992) Changes of the fatty acid composition in smolts of masu salmon (*Oncorhynchus masou*), associated with desmoltification and sea-water transfer. *Comparative Biochemistry and Physiology Part A: Physiology* **103**, 221-6.
- Li H. & Durbin R. (2009) Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics* **25**, 1754-60.
- Li L. & Guo X. (2004) AFLP-based genetic linkage maps of the pacific oyster *Crassostrea gigas* thunberg. *Mar Biotechnol* **6**.
- Li L., Xiang J., Liu X., Zhang Y., Dong B. & Zhang X. (2005) Construction of AFLP-based genetic linkage map for zhikong scallop, *Chlamys farreri* Jones et Preston and mapping of sex-linked markers. *Aquaculture* **245**, 63-73.
- Li M., Sun Y., Zhao J., Shi H., Zeng S., Ye K., Jiang D., Zhou L., Sun L., Tao W., Nagahama Y., Kocher T.D. & Wang D. (2015a) A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PloS Genetics* **11**, e1005678.
- Li W.-H., Gojobori T. & Nei M. (1981) Pseudogenes as a paradigm of neutral evolution. *Nature* **292**, 237-9.
- Li X.H., Bai J.J., Ye X., Hu Y.C., Li S.J. & Yu L.Y. (2009) Polymorphisms in the 5' flanking region of the insulin-like growth factor I gene are associated with growth traits in largemouth bass *Micropterus salmoides*. *Fisheries Science* **75**, 351-8.
- Li X.H., Wu J.S., Tang L.H. & Hu D. (2015b) Identification of conserved microRNAs and their target genes in Nile tilapia (*Oreochromis niloticus*) by bioinformatic analysis. *Genetics and Molecular Research* **14**, 2785-92.
- Li Y., Byrne K., Miggiano E., Whan V., Moore S., Keys S., Crocos P., Preston N. & Lehnert S. (2003) Genetic mapping of the kuruma prawn *Penaeus japonicus* using AFLP markers. *Aquaculture* **219**, 143-56.

- Li Y., Liu S., Qin Z., Waldbieser G., Wang R., Sun L., Bao L., Danzmann R.G., Dunham R. & Liu Z. (2015c) Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. *DNA Research: An International Journal for Rapid Publication of Reports on Genes and Genomes* **22**, 39-52.
- Li Y., Monroig O., Zhang L., Wang S., Zheng X., Dick J.R., You C. & Tocher D.R. (2010) Vertebrate fatty acyl desaturase with $\Delta 4$ activity. *Proceedings of the National Academy of Sciences of the United States of America* **107**, 16840-5.
- Liao X., Cheng L., Xu P., Lu G., Wachholtz M., Sun X. & Chen S. (2013) Transcriptome analysis of Crucian carp (*Carassius auratus*), an important aquaculture and hypoxia-tolerant species. *Plos One* **8**, e62308.
- Likongwe J.S., Stecko T.D., Stauffer J.R. & Carline R.F. (1996) Combined effects of water temperature and salinity on growth and feed utilization of juvenile Nile tilapia *Oreochromis niloticus* (Linnaeus). *Aquaculture* **146**, 37-46.
- Lin G., Chang A., Yap H.W. & Yue G.H. (2008) Characterization and cross-species amplification of microsatellites from the endangered Hawksbill turtle (*Eretmochelys imbricate*). *Conservation Genetics* **9**, 1071-3.
- Lin G., Feng F. & Yue G.H. (2007) Isolation and characterization of polymorphic microsatellites from Asian green mussel (*Perna viridis*). *Molecular Ecology Notes* **7**, 1036-8.
- Lind C.E., Safari A., Agyakwah S.K., Attipoe F.Y.K., El-Naggar G.O., Hamzah A., Hulata G., Ibrahim N.A., Khaw H.L., Nguyen N.H., Maluwa A.O., Zaid M., Zak T. & Ponzoni R.W. (2015) Differences in sexual size dimorphism among farmed tilapia species and strains undergoing genetic improvement for body weight. *Aqua Rep* **1**, 20-7.
- Lindberg J. & Lundberg J. (2010) The plasticity of the mammalian transcriptome. *Genomics* **95**, 1-6.
- Liu B.H. (1997) *Statistical genomics: linkage, mapping, and QTL analysis*. CRC press LLC, U.S.A., New York.
- Liu F., Sun F., Li J., Xia J.H., Lin G., Tu R.J. & Yue G.H. (2013a) A microsatellite-based linkage map of salt tolerant tilapia (*Oreochromis mossambicus* x *Oreochromis spp.*) and mapping of sex-determining loci. *BMC Genomics* **14**, 58.
- Liu F., Sun F., Xia J.H., Li J., Fu G.H., Lin G., Tu R.J., Wan Z.Y., Quek D. & Yue G.H. (2014a) A genome scan revealed significant associations of growth traits with a major QTL and GHR2 in tilapia. *Scientific Reports* **4**, 7256.
- Liu J., Shikano T., Leinonen T., Cano J.M., Li M.-H. & Merilä J. (2014b) Identification of major and minor QTL for ecologically important morphological traits in Three-Spined Sticklebacks (*Gasterosteus aculeatus*). *G3: Genes/Genomes/Genetics* **4**, 595-604.
- Liu L., Yu X. & Tong J. (2012a) Molecular characterization of myostatin (MSTN) gene and association analysis with growth traits in the bighead carp (*Aristichthys nobilis*). *Molecular Biology Reports* **39**, 9211-21.
- Liu S., Sun L., Li Y., Sun F., Jiang Y., Zhang Y., Zhang J., Feng J., Kaltenboeck L., Kucuktas H. & Liu Z. (2014c) Development of the catfish 250K SNP array for genome-wide association studies. *BMC Research Notes* **7**, 135.
- Liu S., Wang X., Sun F., Zhang J., Feng J., Liu H., Rajendran K.V., Sun L., Zhang Y., Jiang Y., Peatman E., Kaltenboeck L., Kucuktas H. & Liu Z. (2013b) RNA-Seq reveals expression signatures of genes involved in oxygen transport, protein synthesis, folding, and degradation in response to heat stress in catfish. *Physiological Genomics* **45**, 462-76.
- Liu S., Zhang Y., Sun F., Jiang Y., Wang R., Li C., Zhang J. & Liu Z. (2012b) Functional genomics research in aquaculture: Principles and general approaches. In: *Functional Genomics in Aquaculture* (eds. by Saroglia M & Liu Z), pp. 1-40. Wiley-Blackwell, Ames, IA.
- Liu S.K., Zhou Z.C., Lu J.G., Sun F.Y., Wang S.L., Liu H., Jiang Y.L., Kucuktas H., Kaltenboeck L., Peatman E. & Liu Z.J. (2011) Generation of genome-scale gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics* **12**.
- Liu Z.-F., Gao X.-Q., Yu J.-X., Qian X.-M., Xue G.-P., Zhang Q.-Y., Liu B.-L. & Hong L. (2017) Effects of different salinities on growth performance, survival, digestive enzyme activity, immune response, and muscle fatty acid composition in juvenile American shad (*Alosa sapidissima*). *Fish Physiology and Biochemistry* **43**, 761-73.
- Liu Z., Liu S., Yao J., Bao L., Zhang J., Li Y., Jiang C., Sun L., Wang R., Zhang Y., Zhou T., Zeng Q., Fu Q., Gao S., Li N., Koren S., Jiang Y., Zimin A., Xu P., Phillippy A.M., Geng X., Song L., Sun F., Li C., Wang X., Chen A., Jin Y., Yuan Z., Yang Y., Tan S., Peatman E., Lu J., Qin Z., Dunham R., Li Z., Sonstegard T., Feng J., Danzmann R.G., Schroeder S., Scheffler B., Duke M.V., Ballard L., Kucuktas H., Kaltenboeck L., Liu H., Armbruster J., Xie Y., Kirby M.L., Tian Y., Flanagan M.E., Mu W. & Waldbieser G.C. (2016) The channel catfish genome sequence provides insights into the evolution of scale formation in teleosts. *Nature Communications* **7**, 11757.
- Liu Z.J. (2007) Fish genomics and analytical genetic technologies, with examples of their potential applications in management of fish genetic resources. *America* **5**, 145-79.
- Liu Z.J. & Cordes J.F. (2004) DNA marker technologies and their applications in aquaculture genetics. *Aquaculture* **238**, 1-37.

- Liu Z.J., Karsi A., Li P., Cao D.F. & Dunham R. (2003) An AFLP-based genetic linkage map of channel catfish (*Ictalurus punctatus*) constructed by using an interspecific hybrid resource family. *Genetics* **165**, 687-94.
- Livak K.J. & Schmittgen T.D. (2001) Analysis of relative gene expression data using real-time quantitative PCR and the 2- $\Delta\Delta$ CT method. *Methods* **25**, 402-8.
- Lorch P.D. (2005) Sex differences in recombination and mapping adaptations. *Genetica* **123**, 39-47.
- Lorenzen K. (2000) Population dynamics and management. In: *Tilapias: Biology and Exploitation* (eds. by Beveridge MCM & McAndrew BJ), pp. 163-225. Springer Netherlands, Dordrecht.
- Louro B., Kuhl H., Tine M., de Koning D.-J., Batargias C., Volckaert F.A.M., Reinhardt R., Canario A.V.M. & Power D.M. (2016) Characterization and refinement of growth related quantitative trait loci in European sea bass (*Dicentrarchus labrax*) using a comparative approach. *Aquaculture* **455**, 8-21.
- Luo z., Liu Y.-J., Mai K.-S., Tian L.-X., Liu D.-H., Tan X.-Y. & Lin H.-Z. (2005) Effect of dietary lipid level on growth performance, feed utilization and body composition of grouper *Epinephelus coioides* juveniles fed isonitrogenous diets in floating netcages. *Aquaculture International* **13**, 257-69.
- Lv W., Zheng X., Kuang Y., Cao D., Yan Y. & Sun X. (2016) QTL variations for growth-related traits in eight distinct families of common carp (*Cyprinus carpio*). *BMC Genetics* **17**, 65.
- Ma H., Chen S., Yang J., Chen S. & Liu H. (2011a) Genetic linkage maps of barfin flounder (*Verasper moseri*) and spotted halibut (*Verasper variegatus*) based on AFLP and microsatellite markers. *Molecular Biology Reports* **38**, 4749-64.
- Ma K., Liao M., Liu F., Ye B., Sun F. & Yue G.H. (2016a) Characterizing the ZFAND3 gene mapped in the sex-determining locus in hybrid tilapia (*Oreochromis spp.*). *Scientific Reports* **6**, 25471.
- Ma Q., Liu S.-F., Zhuang Z.-M., Sun Z.-Z., Liu C.-L., Su Y.-Q. & Tang Q.-S. (2011b) Molecular cloning, expression analysis of insulin-like growth factor I (IGF-I) gene and IGF-I serum concentration in female and male Tongue sole (*Cynoglossus semilaevis*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **160**, 208-14.
- Ma X., Dai W., Kang J., Yang L. & He S. (2016b) Comprehensive transcriptome analysis of six catfish species from an altitude gradient reveals adaptive evolution in Tibetan fishes. *G3: Genes/Genomes/Genetics* **6**, 141-8.
- Macintosh D.J. & Little D.C. (1995) Nile Tilapia (*Oreochromis Niloticus*). In: *Broodstock management and egg and larval quality* (eds. by Bromage NR & Roberts RJ), pp. 277-320. Blackwell, Cambridge.
- Magnoni L.J., Vraskou Y., Palstra A.P. & Planas J.V. (2012) AMP-Activated protein kinase plays an important evolutionary conserved role in the regulation of glucose metabolism in fish skeletal muscle cells. *Plos One* **7**, e31219.
- Mair G.C., Abucay J.S., Beardmore J.A. & Skibinski D.O.F. (1995) Growth performance trials of genetically male tilapia (GMT) derived from YY-males in *Oreochromis niloticus* L.: On station comparisons with mixed sex and sex reversed male populations. *Aquaculture* **137**, 313-23.
- Majumdar K.C. & McAndrew B.J. (1986) Relative DNA content of somatic nuclei and chromosomal studies in three genera, Tilapia, Sarotherodon, and Oreochromis of the tribe Tilapiini (*Pisces, Cichlidae*). *Genetica* **68**, 175-88.
- Manor M.L., Cleveland B.M., Brett Kenney P., Yao J. & Leeds T. (2015) Differences in growth, fillet quality, and fatty acid metabolism-related gene expression between juvenile male and female rainbow trout. *Fish Physiology and Biochemistry* **41**, 533-47.
- Mao X., Young B.D. & Lu Y.-J. (2007) The application of single nucleotide polymorphism microarrays in cancer research. *Current Genomics* **8**, 219-28.
- Marchini J. & Howie B. (2010) Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**, 499-511.
- Mardis E.R. (2006) Anticipating the \$1,000 genome. *Genome Biology* **7**, 112-.
- Marioni J.C., Mason C.E., Mane S.M., Stephens M. & Gilad Y. (2008) RNA-seq: An assessment of technical reproducibility and comparison with gene expression arrays. *Genome Research* **18**, 1509-17.
- Martínez O. & Curnow R.N. (1992) Estimating the locations and the sizes of the effects of quantitative trait loci using flanking markers. *Theoretical and Applied Genetics* **85**, 480-8.
- Martins C., Oliveira C., Wasko A.P. & Wright J.M. (2004) Physical mapping of the Nile tilapia (*Oreochromis niloticus*) genome by fluorescent in situ hybridization of repetitive DNAs to metaphase chromosomes—a review. *Aquaculture* **231**, 37-49.
- Matoulova E., Michalova E., Vojtesek B. & Hrstka R. (2012) The role of the 3' untranslated region in post-transcriptional regulation of protein expression in mammalian cells. *RNA Biology* **9**, 563-76.
- Matsumura H., Ito A., Saitoh H., Winter P., Kahl G., Reuter M., Krüger D.H. & Terauchi R. (2005) SuperSAGE. *Cellular Microbiology* **7**, 11-8.
- McAndrew B. & Napier J. (2011) Application of genetics and genomics to aquaculture development: current and future directions. *The Journal of Agricultural Science* **149**, 143-51.
- McAndrew B.J., Penman D.J., Bekaert M. & Wehner S. (2016) 5 - Tilapia genomic studies A2 - MacKenzie, Simon. In: *Genomics in Aquaculture* (ed. by Jentoft S), pp. 105-29. Academic Press, San Diego.
- McMenamin S.K., Minchin J.E., Gordon T.N., Rawls J.F. & Parichy D.M. (2013) Dwarfism and increased adiposity in the gh1 mutant zebrafish vizzini. *Endocrinology* **154**, 1476-87.

- Mennigen J.A. (2016) Micromanaging metabolism—a role for miRNAs in teleost energy metabolism. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **199**, 115-25.
- Merino G.E., Piedrahita R.H. & Conklin D.E. (2007) The effect of fish stocking density on the growth of California halibut (*Paralichthys californicus*) juveniles. *Aquaculture* **265**, 176-86.
- Metzker M.L. (2010) Sequencing technologies - the next generation. *Nature Review Genetics* **11**, 31-46.
- Meuwissen T. & Goddard M. (1996) The use of marker haplotypes in animal breeding schemes. *Genetics Selection Evolution* **28**, 161.
- Meuwissen T.H., Hayes B.J. & Goddard M.E. (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* **157**, 1819-29.
- Mignone F., Gissi C., Liuni S. & Pesole G. (2002) Untranslated regions of mRNAs. *Genome Biology* **3**, reviews0004.1-reviews.10.
- Miles C. & Wayne M. (2008) Quantitative trait locus (QTL) analysis. *Nature Education* **1**, 208.
- Miller C.T., Glazer A.M., Summers B.R., Blackman B.K., Norman A.R., Shapiro M.D., Cole B.L., Peichel C.L., Schluter D. & Kingsley D.M. (2014) Modular Skeletal Evolution in Sticklebacks Is Controlled by Additive and Clustered Quantitative Trait Loci. *Genetics* **197**, 405-20.
- Miller M.R., Dunham J.P., Amores A., Cresko W.A. & Johnson E.A. (2007) Rapid and cost-effective polymorphism identification and genotyping using restriction site associated DNA (RAD) markers. *Genome Research* **17**, 240-8.
- Moav R. & Wohlfarth G. (1976) Two-way selection for growth rate in the common carp (*Cyprinus carpio* L.). *Genetics* **82**, 83.
- Moen T., Delghandi M., Wesmajervi M.S., Westgaard J.I. & Fjalestad K.T. (2009) A SNP/microsatellite genetic linkage map of the Atlantic cod (*Gadus morhua*). *Animal Genetics* **40**, 993-6.
- Moen T., Hoyheim B., Munck H. & Gomez-Raya L. (2004) A linkage map of Atlantic salmon (*Salmo salar*) reveals an uncommonly large difference in recombination rate between the sexes. *Animal Genetics* **35**, 81-92.
- Mohan M., Nair S., Bhagwat A., Krishna T.G., Yano M., Bhatia C.R. & Sasaki T. (1997) Genome mapping, molecular markers and marker-assisted selection in crop plants. *Molecular Breeding* **3**, 87-103.
- Mohd-Yusof N.Y., Monroig O., Mohd-Adnan A., Wan K.L. & Tocher D.R. (2010) Investigation of highly unsaturated fatty acid metabolism in the Asian sea bass, *Lates calcarifer*. *Fish Physiology and Biochemistry* **36**, 827-43.
- Monroig Ó., Li Y. & Tocher D.R. (2011) Delta-8 desaturation activity varies among fatty acyl desaturases of teleost fish: High activity in delta-6 desaturases of marine species. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **159**, 206-13.
- Morais S., Mourente G., Ortega A., Tocher J.A. & Tocher D.R. (2011) Expression of fatty acyl desaturase and elongase genes, and evolution of DHA:EPA ratio during development of unfed larvae of Atlantic bluefin tuna (*Thunnus thynnus* L.). *Aquaculture* **313**, 129-39.
- Morais S., Taggart J.B., Guy D.R., Bell J.G. & Tocher D.R. (2012) Hepatic transcriptome analysis of inter-family variability in flesh n-3 long-chain polyunsaturated fatty acid content in Atlantic salmon. *BMC Genomics* **13**, 410.
- Moriya Y., Itoh M., Okuda S., Yoshizawa A.C. & Kanehisa M. (2007) KAAS: an automatic genome annotation and pathway reconstruction server. *Nucleic Acids Research* **35**, W182-W5.
- Mozaffarian D. & Wu J.H.Y. (2012) (n-3) Fatty acids and cardiovascular health: Are effects of EPA and DHA shared or complementary? *The Journal of Nutrition* **142**, 614S-25S.
- Mráz J. & Pickova J. (2008) Differences between lipid content and composition of different parts of fillets from crossbred farmed carp (*Cyprinus carpio*). *Fish Physiology and Biochemistry* **35**, 615.
- Nakamura M.T., Cheon Y., Li Y. & Nara T.Y. (2004) Mechanisms of regulation of gene expression by fatty acids. *Lipids* **39**, 1077-83.
- Nakamura M.T. & Nara T.Y. (2004) Structure, function, and dietary regulation of delta6, delta5, and delta9 desaturases. *Annu Rev Nutr* **24**, 345-76.
- Naylor R.L., Goldburg R.J., Primavera J.H., Kautsky N., Beveridge M.C.M., Clay J., Folke C., Lubchenco J., Mooney H. & Troell M. (2000) Effect of aquaculture on world fish supplies. *Nature* **405**, 1017-24.
- Neira R. (2010) Breeding in aquaculture species: genetic improvement programs in developing countries. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*.
- Nelson D.L., Cox M.M. & Lehninger A.L. (2010) *Lehninger principles of biochemistry*. W.H. Freeman, New York.
- Nelson J.S., Grande T.C. & Wilson M.V.H. (2006) *Fishes of the world*. John Wiley & Sons, Inc.
- Ng W.-K. & Romano N. (2013) A review of the nutrition and feeding management of farmed tilapia throughout the culture cycle. *Reviews in Aquaculture* **5**, 220-54.
- Nguyen N., Ponzoni R., Hamzah A., Yee H., Abu-Bakar K. & Khaw H. (2010a) Genetics of flesh quality in fish. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production (WCGALP)*. WCGALP.
- Nguyen N.H., Ponzoni R.W., Abu-Bakar K.R., Hamzah A., Khaw H.L. & Yee H.Y. (2010b) Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), *Oreochromis niloticus*. *Aquaculture* **305**, 1-5.

- Nguyen N.H., Ponzoni R.W., Yee H.Y., Abu-Bakar K.R., Hamzah A. & Khaw H.L. (2010c) Quantitative genetic basis of fatty acid composition in the GIFT strain of Nile tilapia (*Oreochromis niloticus*) selected for high growth. *Aquaculture* **309**, 66-74.
- Nguyen T.V., Jung H., Nguyen T.M., Hurwood D. & Mather P. (2016) Evaluation of potential candidate genes involved in salinity tolerance in striped catfish (*Pangasianodon hypophthalmus*) using an RNA-Seq approach. *Marine Genomics* **25**, 75-88.
- Nielsen E.E., Hemmer-Hansen J., Larsen P.F. & Bekkevold D. (2009) Population genomics of marine fishes: identifying adaptive variation in space and time. *Molecular Ecology* **18**, 3128-50.
- Nyholt D.R. (2000) All LODs are not created equal. *American Journal of Human Genetics* **67**, 282-8.
- O'Malley K.G., Sakamoto T., Danzmann R.G. & Ferguson M.M. (2003) Quantitative trait loci for spawning date and body weight in rainbow trout: testing for conserved effects across ancestrally duplicated chromosomes. *Journal of Heredity* **94**, 273-84.
- O'Fallon J.V., Busboom J.R., Nelson M.L. & Gaskins C.T. (2007) A direct method for fatty acid methyl ester synthesis: Application to wet meat tissues, oils, and feedstuffs. *Journal of Animal Science* **85**, 1511-21.
- Ogden R., Gharbi K., Mugue N., Martinsohn J., Senn H., Davey J.W., Pourkazemi M., McEwing R., Eland C., Vidotto M., Sergeev A. & Congiu L. (2013) Sturgeon conservation genomics: SNP discovery and validation using RAD sequencing. *Molecular Ecology* **22**, 3112-23.
- Olsson G.B., Olsen R.L., Carlehog M. & Ofstad R. (2003) Seasonal variations in chemical and sensory characteristics of farmed and wild Atlantic halibut (*Hippoglossus hippoglossus*). *Aquaculture* **217**, 191-205.
- Ordoñez J.F.F., Ventolero M.F.H. & Santos M.D. (2016) Maternal mismatches in farmed tilapia strains (*Oreochromis spp.*) in the Philippines as revealed by mitochondrial COI gene. *Mitochondrial DNA Part A* **28**, 526-35.
- Ozsolak F. & Milos P.M. (2011) RNA sequencing: advances, challenges and opportunities. *Nat Rev Genet* **12**, 87-98.
- Palaiokostas C., Bekaert M., Khan M.G., Taggart J.B., Gharbi K., McAndrew B.J. & Penman D.J. (2013) Mapping and validation of the major sex-determining region in Nile tilapia (*Oreochromis niloticus* L.) Using RAD sequencing. *Plos One* **8**, e68389.
- Palaiokostas C., Bekaert M., Taggart J.B., Gharbi K., McAndrew B.J., Chatain B., Penman D.J. & Vandeputte M. (2015) A new SNP-based vision of the genetics of sex determination in European sea bass (*Dicentrarchus labrax*). *Genetics Selection Evolution* **47**, 68.
- Palaiokostas C., Ferrarese S., Franch R., Houston R.D. & Bargelloni L. (2016) Genomic prediction of resistance to pasteurellosis in gilthead sea bream (*Sparus aurata*) using 2b-RAD sequencing. *G3: Genes/Genomes/Genetics*.
- Palti Y., Gao G., Liu S., Kent M.P., Lien S., Miller M.R., Rexroad C.E. & Moen T. (2015a) The development and characterization of a 57K single nucleotide polymorphism array for rainbow trout. *Molecular Ecology Resources* **15**, 662-72.
- Palti Y., Vallejo R.L., Gao G., Liu S., Hernandez A.G., Rexroad C.E. & Wiens G.D. (2015b) Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *Plos One* **10**, e0138435.
- Pascual M. & Pla M. (2007) Changes in carcass composition and meat quality when selecting rabbits for growth rate. *Meat Science* **77**, 474-81.
- Pasquier J., Cabau C., Nguyen T., Jouanno E., Severac D., Braasch I., Journot L., Pontarotti P., Klopp C., Postlethwait J.H., Guiguen Y. & Bobe J. (2016) Gene evolution and gene expression after whole genome duplication in fish: the PhyloFish database. *BMC Genomics* **17**, 368.
- Patel R.K. & Jain M. (2012) NGS QC toolkit: A toolkit for quality control of next generation sequencing data. *Plos One* **7**, e30619.
- Paterson A.H. (1997) *Molecular dissection of complex traits*. CRC press.
- Paterson A.H., Lander E.S., Hewitt J.D., Peterson S., Lincoln S.E. & Tanksley S.D. (1988) Resolution of quantitative traits into Mendelian factors by using a complete linkage map of restriction fragment length polymorphisms. *Nature* **335**, 721-6.
- Pelechano V., Wei W. & Steinmetz L.M. (2013) Extensive transcriptional heterogeneity revealed by isoform profiling. *Nature* **497**, 127-31.
- Peng M., Xu W., Mai K., Zhou H., Zhang Y., Liufu Z., Zhang K. & Ai Q. (2014) Growth performance, lipid deposition and hepatic lipid metabolism related gene expression in juvenile turbot (*Scophthalmus maximus* L.) fed diets with various fish oil substitution levels by soybean oil. *Aquaculture* **433**, 442-9.
- Peng X., Li F., Lin S. & Chen Y. (2016) Effects of total replacement of fish oil on growth performance, lipid metabolism and antioxidant capacity in tilapia (*Oreochromis niloticus*). *Aquaculture International* **24**, 145-56.
- Peterson B.C., Waldbieser G.C. & Bilodeau L. (2004) IGF-I and IGF-II mRNA expression in slow and fast growing families of USDA103 channel catfish (*Ictalurus punctatus*). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology* **139**, 317-23.
- Peterson B.K., Weber J.N., Kay E.H., Fisher H.S. & Hoekstra H.E. (2012) Double digest RADseq: an inexpensive method for de novo SNP discovery and genotyping in model and non-model species. *Plos One* **7**, e37135.

- Pfennig F., Standke A. & Gutzeit H.O. (2015) The role of Amh signaling in teleost fish – Multiple functions not restricted to the gonads. *General and Comparative Endocrinology* **223**, 87-107.
- Phillips P.C. & Kohler C.C. (1991) Establishment of Tilapia spawning families providing a continuous supply of eggs for in vitro fertilization. *Journal of the World Aquaculture Society* **22**, 217-23.
- Polakof S., Panserat S., Craig P.M., Martyres D.J., Plagnes-Juan E., Savari S., Aris-Brosou S. & Moon T.W. (2011) The metabolic consequences of hepatic AMP-Kinase phosphorylation in rainbow trout. *Plos One* **6**, e20228.
- Poland J.A., Brown P.J., Sorrells M.E. & Jannink J.-L. (2012) Development of high-density genetic maps for barley and wheat using a novel two-enzyme genotyping-by-sequencing approach. *Plos One* **7**, e32253.
- Poland J.A. & Rife T.W. (2012) Genotyping-by-sequencing for plant breeding and genetics. *Plant Genome-U.S.* **5**.
- Pompanon F., Bonin A., Bellemain E. & Taberlet P. (2005) Genotyping errors: causes, consequences and solutions. *Nat Rev Genet* **6**, 847-6.
- Pongthana N., Nguyen N.H. & Ponzoni R.W. (2010) Comparative performance of four red tilapia strains and their crosses in fresh- and saline water environments. *Aquaculture* **308**, S109-S14.
- Ponzoni R.W., Nguyen N.H., Khaw H.L., Hamzah A., Bakar K.R.A. & Yee H.Y. (2011) Genetic improvement of Nile tilapia (*Oreochromis niloticus*) with special reference to the work conducted by the WorldFish Center with the GIFT strain. *Reviews in Aquaculture* **3**, 27-41.
- Porazinski S., Wang H., Asaoka Y., Behrndt M., Miyamoto T., Morita H., Hata S., Sasaki T., Krens S.F.G., Osada Y., Asaka S., Momoi A., Linton S., Miesfeld J.B., Link B.A., Senga T., Castillo-Morales A., Urrutia A.O., Shimizu N., Nagase H., Matsuura S., Bagby S., Kondoh H., Nishina H., Heisenberg C.-P. & Furutani-Seiki M. (2015) YAP is essential for tissue tension to ensure vertebrate 3D body shape. *Nature* **521**, 217-21.
- Powell J., White I., Guy D. & Brotherstone S. (2008) Genetic parameters of production traits in Atlantic salmon (*Salmo salar*). *Aquaculture* **274**, 225-31.
- Pozios K.C., Ding J., Degger B., Upton Z. & Duan C. (2001) IGFs stimulate zebrafish cell proliferation by activating MAP kinase and PI3-kinase-signaling pathways. *American Journal of Physiology - Regulatory, Integrative and Comparative Physiology* **280**, R1230.
- Preziosa E., Liu S., Terova G., Gao X., Liu H., Kucuktas H., Terhune J. & Liu Z. (2013) Effect of nutrient restriction and re-feeding on Calpain family genes in skeletal muscle of Channel catfish (*Ictalurus punctatus*). *Plos One* **8**, e59404.
- Qin Y., Liu X., Zhang H., Zhang G. & Guo X. (2007) Genetic mapping of size-related quantitative trait loci (QTL) in the bay scallop (*Argopecten irradians*) using AFLP and microsatellite markers. *Aquaculture* **272**, 281-90.
- Quail M.A., Smith M., Coupland P., Otto T.D., Harris S.R., Connor T.R., Bertoni A., Swerdlow H.P. & Gu Y. (2012) A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics* **13**, 341-.
- Quillet E., Le Guillou S., Aubin J. & Fauconneau B. (2005) Two-way selection for muscle lipid content in pan-size rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* **245**, 49-61.
- Quinton C.D., McMillan I. & Glebe B.D. (2005) Development of an Atlantic salmon (*Salmo salar*) genetic improvement program: Genetic parameters of harvest body weight and carcass quality traits estimated with animal models. *Aquaculture* **247**, 211-7.
- Rahnan S.A., Huah T.S., Nassan O. & Daud N.M. (1995) Fatty acid composition of some Malaysian freshwater fish. *Food Chemistry* **54**, 45-9.
- Rasoarahona J.R.E., Barnathan G., Bianchini J.-P. & Gaydou E.M. (2005) Influence of season on the lipid content and fatty acid profiles of three tilapia species (*Oreochromis niloticus*, *O. macrochir* and *Tilapia rendalli*) from Madagascar. *Food Chemistry* **91**, 683-94.
- Recknagel H., Elmer K.R. & Meyer A. (2013) A hybrid genetic linkage map of two ecologically and morphologically divergent Midas Cichlid fishes (*Amphilophus* spp.) obtained by massively parallel DNA sequencing (ddRADSeq). *G3-Genes Genomes Genetics* **3**.
- Reed T.E., Prodohl P., Hynes R., Cross T., Ferguson A. & McGinnity P. (2015) Quantifying heritable variation in fitness-related traits of wild, farmed and hybrid Atlantic salmon families in a wild river environment. *Heredity* **115**, 173-84.
- Reid D.P., Smith C.-A., Rommens M., Blanchard B., Martin-Robichaud D. & Reith M. (2007) A genetic linkage map of Atlantic halibut (*Hippoglossus hippoglossus* L.). *Genetics* **177**, 1193-205.
- Reid D.P., Szanto A., Glebe B., Danzmann R.G. & Ferguson M.M. (2004) QTL for body weight and condition factor in Atlantic salmon (*Salmo salar*): comparative analysis with rainbow trout (*Oncorhynchus mykiss*) and Arctic charr (*Salvelinus alpinus*). *Heredity* **94**, 166-72.
- Reinecke M. (2010) Influences of the environment on the endocrine and paracrine fish growth hormone–insulin-like growth factor-I system. *J. Fish Biol* **76**, 1233-54.
- Ren H.-t., Yu J.-h., Xu P. & Tang Y.-k. (2012) Influence of dietary fatty acids on muscle fatty acid composition and expression levels of $\Delta 6$ desaturase-like and Elovl5-like elongase in common carp (*Cyprinus carpio* var. *Jian*). *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **163**, 184-92.
- Rengmark A.H. & Lingaas F. (2007) Genomic structure of the Nile tilapia (*Oreochromis niloticus*) transferrin gene and a haplotype associated with saltwater tolerance. *Aquaculture* **272**, 146-55.

- Rengmark A.H., Slettan A., Lee W.J., Lie Ø. & Lingaas F. (2007) Identification and mapping of genes associated with salt tolerance in tilapia. *Journal of Fish Biology* **71**, 409-22.
- Rescan P.-Y., Montfort J., Rallière C., Le Cam A., Esquerré D. & Hugot K. (2007) Dynamic gene expression in fish muscle during recovery growth induced by a fasting-refeeding schedule. *BMC Genomics* **8**, 1.
- Rexroad C.E., Palti Y., Gahr S.A. & Vallejo R.L. (2008) A second generation genetic map for rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **9**, 74.
- Ribaut J.-M. & Ragot M. (2007) Marker-assisted selection to improve drought adaptation in maize: the backcross approach, perspectives, limitations, and alternatives. *Journal of Experimental Botany* **58**, 351-60.
- Rimoldi S., Benedito-Palos L., Terova G. & Pérez-Sánchez J. (2016) Wide-targeted gene expression infers tissue-specific molecular signatures of lipid metabolism in fed and fasted fish. *Reviews in Fish Biology and Fisheries* **26**, 93-108.
- Risch N. (1992) Genetic linkage: interpreting lod scores. *Science* **255**, 803-4.
- Robb D.H.F., Kestin S.C., Warriss P.D. & Nute G.R. (2002) Muscle lipid content determines the eating quality of smoked and cooked Atlantic salmon (*Salmo salar*). *Aquaculture* **205**, 345-58.
- Robledo D., Fernández C., Hermida M., Sciara A., Álvarez-Dios J.A., Cabaleiro S., Caamaño R., Martínez P. & Bouza C. (2016) Integrative transcriptome, genome and quantitative trait loci resources identify single nucleotide polymorphisms in candidate genes for growth traits in Turbot. *International Journal of Molecular Sciences* **17**, 243.
- Robledo D., Palaikostas C., Bargelloni L., Martínez P. & Houston R. (2017) Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in Aquaculture*, doi:10.1111/raq.12193.
- Roch G.J., Wu S. & Sherwood N.M. (2009) Hormones and receptors in fish: do duplicates matter? *General and Comparative Endocrinology* **161**, 3-12.
- Romana-Eguia M.R.R. & Eguia R.V. (1999) Growth of five Asian red tilapia strains in saline environments. *Aquaculture* **173**, 161-70.
- Ross J.A., Urton J.R., Boland J., Shapiro M.D. & Peichel C.L. (2009) Turnover of sex chromosomes in the stickleback fishes (*Gasterosteidae*). *PloS Genetics* **5**, e1000391.
- Rothschild M.F. & Ruvinsky A. (2007) Marker-assisted selection for aquaculture species. In: *Aquaculture Genome Technologies* (ed. by Liu ZJ), pp. 199-214. Blackwell Publishing Ltd.
- Rowlerson A. & Veggetti A. (2001) Cellular mechanisms of post embryonic muscle growth in aquaculture species. In: *Fish physiology 18: Muscle development and growth*, pp. 103-40.
- Ruan X., Wang W., Kong J., Yu F. & Huang X. (2010) Genetic linkage mapping of turbot (*Scophthalmus maximus* L.) using microsatellite markers and its application in QTL analysis. *Aquaculture* **308**, 89-100.
- Rutten M.J.M., Bovenhuis H. & Komen H. (2005) Genetic parameters for fillet traits and body measurements in Nile tilapia (*Oreochromis niloticus* L.). *Aquaculture* **246**, 125-32.
- Rye M. & Gjerde B. (1996) Phenotypic and genetic parameters of body composition traits and flesh colour in Atlantic salmon, *Salmo salar* L. *Aquaculture Research* **27**, 121-33.
- Sachidanandam R., Weissman D., Schmidt S.C., Kakol J.M., Stein L.D., Marth G., Sherry S., Mullikin J.C., Mortimore B.J. & Willey D.L. (2001) A map of human genome sequence variation containing 1.42 million single nucleotide polymorphisms. *Nature* **409**, 928-33.
- Saillant E., Dupont-Nivet M., Sabourault M., Haffray P., Laureau S., Vidal M.-O. & Chatain B. (2009) Genetic variation for carcass quality traits in cultured sea bass (*Dicentrarchus labrax*). *Aquatic Living Resources* **22**, 105-12.
- Sakamoto T., Danzmann R.G., Gharbi K., Howard P., Ozaki A., Khoo S.K., Woram R.A., Okamoto N., Ferguson M.M., Holm L.-E., Guyomard R. & Hoyheim B. (2000) A microsatellite linkage map of rainbow trout (*Oncorhynchus mykiss*) characterized by large sex-specific differences in recombination rates. *Genetics* **155**, 1331-45.
- Salem M., Vallejo R.L., Leeds T.D., Palti Y., Liu S., Sabbagh A., Rexroad C.E., III & Yao J. (2012) RNA-Seq identifies SNP markers for growth traits in Rainbow trout. *Plos One* **7**, e36264.
- Sánchez-Molano E., Cerna A., Toro M.A., Bouza C., Hermida M. & Pardo B.G. (2011) Detection of growth-related QTL in turbot (*Scophthalmus maximus*). *BMC Genomics* **12**, 473.
- Sanchez M.-P., Iannuccelli N., Basso B., Bidanel J.-P., Billon Y., Gandemer G., Gilbert H., Larzul C., Legault C., Riquet J., Milan D. & Le Roy P. (2007) Identification of QTL with effects on intramuscular fat content and fatty acid composition in a Duroc × Large White cross. *BMC Genetics* **8**, 55.
- Sander J.D. & Joung J.K. (2014) CRISPR-Cas systems for editing, regulating and targeting genomes. *Nat Biotech* **32**, 347-55.
- Santos V.B.d., Mareco E.A. & Dal Pai Silva M. (2013) Growth curves of Nile tilapia (*Oreochromis niloticus*) strains cultivated at different temperatures. *Acta Scientiarum. Animal Sciences* **35**, 235-42.
- Santulli A. & D'Amelio V. (1986) Effects of supplemental dietary carnitine on growth and lipid metabolism of hatchery-reared sea bass (*Dicentrarchus labrax* L.). *Aquaculture* **59**, 177-86.
- Sargent J., Bell G., McEvoy L., Tocher D. & Estevez A. (1999) Recent developments in the essential fatty acid nutrition of fish. *Aquaculture* **177**, 191-9.

- Sargent J.R. & Tacon A.G.J. (1999) Development of farmed fish: a nutritionally necessary alternative to meat. *Proceedings of the Nutrition Society* **58**, 377-83.
- Sargent J.R., Tocher D.R. & Bell J.G. (2003) 4 - The Lipids A2. In: *Fish Nutrition (Third Edition)* (eds. by Halver JE & Hardy RW), pp. 181-257. Academic Press, San Diego.
- Schmid T., Jansen A.P., Baker A.R., Hegamyer G., Hagan J.P. & Colburn N.H. (2008) Translation inhibitor Pdcd4 is targeted for degradation during tumor promotion. *Cancer Research* **68**, 1254-60.
- Schuler G.D., Boguski M.S., Stewart E.A., Stein L.D., Gyapay G., Rice K., White R.E., Rodriguez-Tomé P., Aggarwal A., Bajorek E., Bentolila S., Birren B.B., Butler A., Castle A.B., Chiannikulchai N., Chu A., Clee C., Cowles S., Day P.J.R., Dibbling T., Drouot N., Dunham I., Duprat S., East C., Edwards C., Fan J.-B., Fang N., Fizames C., Garrett C., Green L., Hadley D., Harris M., Harrison P., Brady S., Hicks A., Holloway E., Hui L., Hussain S., Louis-Dit-Sully C., Ma J., MacGilvery A., Mader C., Maratukulam A., Matisse T.C., McKusick K.B., Morissette J., Mungall A., Muselet D., Nusbaum H.C., Page D.C., Peck A., Perkins S., Piercy M., Qin F., Quackenbush J., Ranby S., Reif T., Rozen S., Sanders C., She X., Silva J., Slonim D.K., Soderlund C., Sun W.-L., Tabar P., Thangarajah T., Vega-Czarny N., Vollrath D., Voyticky S., Wilmer T., Wu X., Adams M.D., Auffray C., Walter N.A.R., Brandon R., Dehejia A., Goodfellow P.N., Houlgate R., Hudson J.R., Ide S.E., Iorio K.R., Lee W.Y., Seki N., Nagase T., Ishikawa K., Nomura N., Phillips C., Polymeropoulos M.H., Sandusky M., Schmitt K., Berry R., Swanson K., Torres R., Venter J.C., Sikela J.M., Beckmann J.S., Weissenbach J., Myers R.M., Cox D.R., James M.R., Bentley D., Deloukas P., Lander E.S. & Hudson T.J. (1996) A gene map of the human genome. *Science* **274**, 540-6.
- Schunter C., Vollmer S.V., Macpherson E. & Pascual M. (2014) Transcriptome analyses and differential gene expression in a non-model fish species with alternative mating tactics. *BMC Genomics* **15**, 167.
- Schütz D. & Taborsky M. (2000) Giant males or dwarf females: what determines the extreme sexual size dimorphism in *Lamprologus callipterus*? *Journal of Fish Biology* **57**, 1254-65.
- Sekino M. & Hara M. (2007) Linkage maps for the Pacific abalone (*Genus Haliotis*) Based on microsatellite DNA markers. *Genetics* **175**, 945-58.
- Serhan C.N., Chiang N. & Van Dyke T.E. (2008) Resolving inflammation: dual anti-inflammatory and pro-resolution lipid mediators. *Nature Reviews Immunology* **8**, 349-61.
- Shao C., Li Q., Chen S., Zhang P., Lian J., Hu Q., Sun B., Jin L., Liu S., Wang Z., Zhao H., Jin Z., Liang Z., Li Y., Zheng Q., Zhang Y., Wang J. & Zhang G. (2014) Epigenetic modification and inheritance in sexual reversal of fish. *Genome Research* **24**, 604-15.
- Shao C., Niu Y., Rastas P., Liu Y., Xie Z., Li H., Wang L., Jiang Y., Tai S., Tian Y., Sakamoto T. & Chen S. (2015) Genome-wide SNP identification for the construction of a high-resolution genetic map of Japanese flounder (*Paralichthys olivaceus*): applications to QTL mapping of *Vibrio anguillarum* disease resistance and comparative genomic analysis. *DNA Research* **22**, 161-70.
- Sheehan R.J., Shasteen S.P., Suresh A.V., Kapuscinski A.R. & Seeb J.E. (1999) Better growth in all-female diploid and triploid rainbow trout. *Transactions of the American Fisheries Society* **128**, 491-8.
- Shendure J. & Ji H. (2008) Next-generation DNA sequencing. *Nature Biotechnology* **26**, 1135-45.
- Sheridan M.A. (1988) Lipid dynamics in fish: aspects of absorption, transportation, deposition and mobilization. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **90**, 679-90.
- Sheridan M.A., Allen W.V. & Kerstetter T.H. (1985) Changes in the fatty acid composition of steelhead trout, *Salmo gairdnerii* Richardson, associated with parr-smolt transformation. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **80**, 671-6.
- Shi X., Wallis A.M., Gerard R.D., Voelker K.A., Grange R.W., DePinho R.A., Garry M.G. & Garry D.J. (2012) Foxk1 promotes cell proliferation and represses myogenic differentiation by regulating Foxo4 and Mef2. *Journal of Cell Science* **125**, 5329-37.
- Shi Y., Kui H., Guo X., Gu Z., Wang Y. & Wang A. (2009) Genetic linkage map of the pearl oyster, *Pinctada martensii* (dunker). *Aquaculture Research* **41**, 35-44.
- Shirak A., Palti Y., Cnaani A., Korol A., Hulata G., Ron M. & Avtalion R.R. (2002) Association between loci With deleterious alleles and distorted sex ratios in an inbred line of tilapia (*Oreochromis aureus*). *Journal of Heredity* **93**, 270-6.
- Shirak A., Seroussi E., Cnaani A., Howe A.E., Domokhovskiy R., Zilberman N., Kocher T.D., Hulata G. & Ron M. (2006) Amh and Dmrt2 genes map to tilapia (*Oreochromis spp.*) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* **174**, 1573-81.
- Sidhu K.S. (2003) Health benefits and potential risks related to consumption of fish or fish oil. *Regulatory Toxicology and Pharmacology* **38**, 336-44.
- Sifa L., Chenhong L., Dey M., Galgalac F. & Dunham R. (2002) Cold tolerance of three strains of Nile tilapia, *Oreochromis niloticus*, in China. *Aquaculture* **213**, 123-9.
- Simopoulos A.P. (1999) Evolutionary aspects of omega-3 fatty acids in the food supply. *Prostaglandins, Leukotrienes and Essential Fatty Acids* **60**, 421-9.
- Simopoulos A.P. (2006) Evolutionary aspects of diet, the omega-6/omega-3 ratio and genetic variation: nutritional implications for chronic diseases. *Biomedicine & Pharmacotherapy* **60**, 502-7.
- Sims D., Sudbery I., Illott N.E., Heger A. & Ponting C.P. (2014) Sequencing depth and coverage: key considerations in genomic analyses. *Nat Rev Genet* **15**, 121-32.

- Smith S., Bernatchez L. & Beheregaray L.B. (2013) RNA-seq analysis reveals extensive transcriptional plasticity to temperature stress in a freshwater fish species. *BMC Genomics* **14**, 375.
- Sneller C.H., Mather D.E. & Crepieux S. (2009) Analytical approaches and population types for finding and utilizing QTL in complex plant populations. *Crop Science* **49**, 363-80.
- Sodeland M., Gaarder M., Moen T., Thomassen M., Kjøglum S., Kent M. & Lien S. (2013) Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon. *Aquaculture* **408–409**, 169-74.
- Sonah H., Bastien M., Iqura E., Tardivel A., Legare G., Boyle B., Normandeau E., Laroche J., Larose S., Jean M. & Belzile F. (2013) An improved genotyping by sequencing (GBS) approach offering increased versatility and efficiency of SNP discovery and genotyping. *Plos One* **8**, e54603.
- Sonenberg N., Rupprecht K.M., Hecht S.M. & Shatkin A.J. (1979) Eukaryotic mRNA cap binding protein: purification by affinity chromatography on sepharose-coupled m7GDP. *Proceedings of the National Academy of Sciences* **76**, 4345-9.
- Sonesson A.K. & Meuwissen T.H. (2009) Testing strategies for genomic selection in aquaculture breeding programs. *Genetics Selection Evolution* **41**, 37.
- Song W., Pang R., Niu Y., Gao F., Zhao Y., Zhang J., Sun J., Shao C., Liao X., Wang L., Tian Y. & Chen S. (2012) Construction of high-density genetic linkage maps and mapping of growth-related quantitative trait loci in the Japanese flounder (*Paralichthys olivaceus*). *Plos One* **7**, e50404.
- Sprecher H. (2000) Metabolism of highly unsaturated n-3 and n-6 fatty acids. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1486**, 219-31.
- Star B., Nederbragt A.J., Jentoft S., Grimholt U., Malmstrom M., Gregers T.F., Rounge T.B., Paulsen J., Solbakken M.H., Sharma A., Wetten O.F., Lanzen A., Winer R., Knight J., Vogel J.-H., Aken B., Andersen O., Lagesen K., Tooming-Klunderud A., Edvardsen R.B., Tina K.G., Espelund M., Nepal C., Previti C., Karlens B.O., Moum T., Skage M., Berg P.R., Gjoen T., Kuhl H., Thorsen J., Malde K., Reinhardt R., Du L., Johansen S.D., Searle S., Lien S., Nilsen F., Jonassen I., Omholt S.W., Stenseth N.C. & Jakobsen K.S. (2011) The genome sequence of Atlantic cod reveals a unique immune system. *Nature* **477**, 207-10.
- Stein S.E. & Scott D.R. (1994) Optimization and testing of mass spectral library search algorithms for compound identification. *Journal of the American Society for Mass Spectrometry* **5**, 859-66.
- Stickney R.R. (1986) Tilapia tolerance of saline waters: A Review. *The Progressive Fish-Culturist* **48**, 161-7.
- Sun J.X., Mullikin J.C., Patterson N. & Reich D.E. (2009) Microsatellites are molecular clocks that support accurate inferences about history. *Molecular Biology and Evolution* **26**, 1017-27.
- Sun L., Liu S., Wang R., Jiang Y., Zhang Y., Zhang J., Bao L., Kaltenboeck L., Dunham R., Waldbieser G. & Liu Z. (2014a) Identification and analysis of genome-wide SNPs provide insight into signatures of selection and domestication in channel catfish (*Ictalurus punctatus*). *Plos One* **9**, e109666.
- Sun Y.-L., Jiang D.-N., Zeng S., Hu C.-J., Ye K., Yang C., Yang S.-J., Li M.-H. & Wang D.-S. (2014b) Screening and characterization of sex-linked DNA markers and marker-assisted selection in the Nile tilapia (*Oreochromis niloticus*). *Aquaculture* **433**, 19-27.
- Sun Y., Huang Y., Li X., Baldwin C.C., Zhou Z., Yan Z., Crandall K.A., Zhang Y., Zhao X., Wang M., Wong A., Fang C., Zhang X., Huang H., Lopez J.V., Kilfoyle K., Zhang Y., Ortí G., Venkatesh B. & Shi Q. (2016) Fish-T1K (Transcriptomes of 1,000 Fishes) Project: large-scale transcriptome data for fish evolution studies. *GigaScience* **5**, 18.
- Sun Y., Yu X. & Tong J. (2012) Polymorphisms in myostatin gene and associations with growth traits in the common carp (*Cyprinus carpio* L.). *International Journal of Molecular Sciences* **13**, 14956-61.
- Sundström L.F., Löhmus M. & Devlin R.H. (2005) Selection on increased intrinsic growth rates in coho salmon, *Oncorhynchus kisutch* *Evolution* **59**, 1560-9.
- Suresh A.V. & Lin C.K. (1992) Tilapia culture in saline waters: a review. *Aquaculture* **106**, 201-26.
- Syaifudin M. (2015) Species-specific DNA markers for improving the genetic management of Tilapia. In: *Institute of Aquaculture*, p. 241. University of Stirling, United Kingdom.
- Syvanen A.-C. (2005) Toward genome-wide SNP genotyping. *Nature Genetics* **37**, S5 - S10.
- Tabor H.K., Risch N.J. & Myers R.M. (2002) Candidate-gene approaches for studying complex genetic traits: practical considerations. *Nat Rev Genet* **3**, 391-7.
- Tacchi L., Bickerdike R., Douglas A., Secombes C.J. & Martin S.A.M. (2011) Transcriptomic responses to functional feeds in Atlantic salmon (*Salmo salar*). *Fish & Shellfish Immunology* **31**, 704-15.
- Tang X.-L., Xu M.-J., Li Z.-H., Pan Q. & Fu J.-H. (2013) Effects of vitamin E on expressions of eight microRNAs in the liver of Nile tilapia (*Oreochromis niloticus*). *Fish & Shellfish Immunology* **34**, 1470-5.
- Tanksley S.D. (1993) Mapping polygenes. *Annual review of genetics* **27**, 205-33.
- Tanomman S., Ketudat-Cairns M., Jangprai A. & Boonanutanasarn S. (2013) Characterization of fatty acid delta-6 desaturase gene in Nile tilapia and heterogenous expression in *Saccharomyces cerevisiae*. *Comparative Biochemistry and Physiology Part B: Biochemistry and Molecular Biology* **166**, 148-56.
- Tao W., Yuan J., Zhou L., Sun L., Sun Y., Yang S., Li M., Zeng S., Huang B. & Wang D. (2013) Characterization of gonadal transcriptomes from Nile tilapia (*Oreochromis niloticus*) reveals differentially expressed genes. *Plos One* **8**, e63604.

- Tao W.J. & Boulding E.G. (2003) Associations between single nucleotide polymorphisms in candidate genes and growth rate in Arctic charr (*Salvelinus alpinus* L.). *Heredity* **91**, 60-9.
- Tarazona S., García-Alcalde F., Dopazo J., Ferrer A. & Conesa A. (2011) Differential expression in RNA-seq: A matter of depth. *Genome Research*.
- Tayamen M., Abella T., Reyes R., Danting M.J., Mendoza A., Marquez E., Salguet A., Apaga M. & Gonzales R. (2002) Development of tilapia for saline waters in the Philippines.
- Thevasagayam N., Sridatta P., Jiang J., Tong A., Saju J., Kathiresan P., Kwan H., Ngoh S., Liew W., Kuznetsova I., Shen X., Lok S., Vij S. & Orbán L. (2015) Transcriptome survey of a marine food fish: Asian seabass (*Lates calcarifer*). *Journal of Marine Science and Engineering* **3**, 382.
- Thodesen J., Rye M., Wang Y.-X., Li S.-J., Bentsen H.B., Yazdi M.H. & Gjedrem T. (2013) Genetic improvement of tilapias in China: Genetic parameters and selection responses in growth, survival and external color traits of red tilapia (*Oreochromis spp.*) after four generations of multi-trait selection. *Aquaculture* **416–417**, 354-66.
- Thomson M.J. (2014) High-throughput SNP genotyping to accelerate crop improvement. *Plant Breeding and Biotechnology* **2**, 195-212.
- Tian J.-J., Lu R.-H., Ji H., Sun J., Li C., Liu P., Lei C.-X., Chen L.-Q. & Du Z.-Y. (2015a) Comparative analysis of the hepatopancreas transcriptome of grass carp (*Ctenopharyngodon idellus*) fed with lard oil and fish oil diets. *Gene* **565**, 192-200.
- Tian M., Li Y., Jing J., Mu C., Du H., Dou J., Mao J., Li X., Jiao W., Wang Y., Hu X., Wang S., Wang R. & Bao Z. (2015b) Construction of a high-density genetic map and quantitative trait locus mapping in the sea cucumber *Apostichopus japonicus*. *Scientific Reports* **5**, 14852.
- Tine M., Kuhl H., Gagnaire P.-A., Louro B., Desmarais E., Martins R.S.T., Hecht J., Knaust F., Belkhir K., Klages S., Dieterich R., Stueber K., Piferrer F., Guinand B., Bierne N., Volckaert F.A.M., Bargelloni L., Power D.M., Bonhomme F., Canario A.V.M. & Reinhardt R. (2014) European sea bass genome and its variation provide insights into adaptation to euryhalinity and speciation. *Nature Communications* **5**, 5770.
- Tobin D., Kause A., Mäntysaari E.A., Martin S.A.M., Houlihan D.F., Dobby A., Kiessling A., Rungruangsak-Torrissen K., Ritola O. & Ruohonen K. (2006) Fat or lean? The quantitative genetic basis for selection strategies of muscle and body composition traits in breeding schemes of rainbow trout (*Oncorhynchus mykiss*). *Aquaculture* **261**, 510-21.
- Tocher D.R. (2003) Metabolism and functions of lipids and fatty acids in teleost fish. *Reviews in Fisheries Science* **11**, 107-84.
- Tocher D.R. & Glencross B.D. (2015) Lipids and fatty acids. In: *Dietary Nutrients, Additives, and Fish Health* (pp. 47-94. John Wiley & Sons, Inc.
- Tocher D.R., Zheng X., Schlechtriem C., Hastings N., Dick J.R. & Teale A.J. (2006) Highly unsaturated fatty acid synthesis in marine fish: Cloning, functional characterization, and nutritional regulation of fatty acyl $\Delta 6$ desaturase of Atlantic cod (*Gadus morhua* L.). *Lipids* **41**, 1003-16.
- Toguyeni A., Fauconneau B., Fostier A., Abucay J., Mair G. & Baroiller J.-F. (2002) Influence of sexual phenotype and genotype, and sex ratio on growth performances in tilapia, *Oreochromis niloticus*. *Aquaculture* **207**, 249-61.
- Torrelee E., Sluiszen A.v.D. & Verreth J. (2007) The effect of dietary L-carnitine on the growth performance in fingerlings of the African catfish (*Clarias gariepinus*) in relation to dietary lipid. *British Journal of Nutrition* **69**, 289-99.
- Tran N.T., Gao Z.-X., Zhao H.-H., Yi S.-K., Chen B.-X., Zhao Y.-H., Lin L., Liu X.-Q. & Wang W.-M. (2015) Transcriptome analysis and microsatellite discovery in the blunt snout bream (*Megalobrama amblycephala*) after challenge with *Aeromonas hydrophila*. *Fish & Shellfish Immunology* **45**, 72-82.
- Trọng T.Q., van Bers N., Crooijmans R., Dibbits B. & Komen H. (2013) A comparison of microsatellites and SNPs in parental assignment in the GIFT strain of Nile tilapia (*Oreochromis niloticus*): The power of exclusion. *Aquaculture* **388–391**, 14-23.
- Tsai H.Y., Hamilton A., Guy D.R. & Houston R.D. (2014) Single nucleotide polymorphisms in the insulin-like growth factor 1 (IGF1) gene are associated with growth-related traits in farmed Atlantic salmon. *Animal Genetics* **45**, 709-15.
- Tsai H.Y., Hamilton A., Guy D.R., Tinch A.E., Bishop S.C. & Houston R.D. (2015) The genetic architecture of growth and fillet traits in farmed Atlantic salmon (*Salmo salar*). *BMC Genetics* **16**, 51.
- Tsai H.Y., Robledo D., Lowe N.R., Bekaert M., Taggart J.B., Bron J.E. & Houston R.D. (2016) Construction and annotation of a high density SNP linkage map of the Atlantic salmon (*Salmo salar*) genome. *G3: Genes/Genomes/Genetics*.
- Uchida K., Kaneko T., Miyazaki H., Hasegawa S. & Hirano T. (2000) Excellent salinity tolerance of Mozambique tilapia (*Oreochromis mossambicus*): Elevated chloride cell activity in the branchial and opercular epithelia of the fish adapted to concentrated Seawater. *Zoological Science* **17**, 149-60.
- Untergasser A., Nijveen H., Rao X., Bisseling T., Geurts R. & Leunissen J.A. (2007) Primer3Plus, an enhanced web interface to Primer3. *Nucleic Acids Research* **35**, W71-W4.
- Valente L.M.P., Moutou K.A., Conceição L.E.C., Engrola S., Fernandes J.M.O. & Johnston I.A. (2013) What determines growth potential and juvenile quality of farmed fish species? *Reviews in Aquaculture* **5**, S168-S93.

- Van de Peer Y., Maere S. & Meyer A. (2009) The evolutionary significance of ancient genome duplications. *Nature Reviews Genetics* **10**, 725-32.
- Van Ooijen J.W. (2004) MapQTL@5.0, Software for the mapping of quantitative trait loci in experimental populations., Kyazma B.V., Wageningen, Netherlands.
- Van Ooijen J.W. (2006) JoinMap 4.0: Software for the calculation of genetic linkage maps in experimental populations. *Wageningen, Netherlands: Kyazma BV*.
- Van Ooijen J.W. (2011) Multipoint maximum likelihood mapping in a full-sib family of an outbreeding species. *Genetic Research* **93**, 343-9.
- Vergara J., eacute, Manuel, Robain, agrave, Lidia, Izquierdo M. & Higuera M.D.L. (1996) Protein sparing effect of lipids in diets for fingerlings of Gilthead sea bream. *Fisheries Science* **62**, 624-8.
- Videler J.J. (2011) An opinion paper: emphasis on white muscle development and growth to improve farmed fish flesh quality. *Fish Physiology and Biochemistry* **37**, 337-43.
- Vieira M.L.C., Santini L., Diniz A.L. & Munhoz C.d.F. (2016) Microsatellite markers: what they mean and why they are so useful. *Genetics and Molecular Biology* **39**, 312-28.
- Vieira V.L.A., Norris A. & Johnston I.A. (2007) Heritability of fibre number and size parameters and their genetic relationship to flesh quality traits in Atlantic salmon (*Salmo salar* L.). *Aquaculture* **272**, S100-S9.
- Vignal A., Milan D., SanCristobal M. & Eggen A. (2002) A review on SNP and other types of molecular markers and their use in animal genetics. *Genetics, Selection, Evolution : GSE* **34**, 275-305.
- Vij S., Kuhl H., Kuznetsova I.S., Komissarov A., Yurchenko A.A., Van Heusden P., Singh S., Thevasagayam N.M., Prakki S.R.S., Purushothaman K., Saju J.M., Jiang J., Mbandi S.K., Jonas M., Hin Yan Tong A., Mwangi S., Lau D., Ngho S.Y., Liew W.C., Shen X., Hon L.S., Drake J.P., Boitano M., Hall R., Chin C.-S., Lachumanan R., Korlach J., Trifonov V., Kabilov M., Tupikin A., Green D., Moxon S., Garvin T., Sedlazeck F.J., Vurture G.W., Gopalapillai G., Kumar Katneni V., Noble T.H., Scaria V., Sivasubbu S., Jerry D.R., O'Brien S.J., Schatz M.C., Dalmay T., Turner S.W., Lok S., Christoffels A. & Orbán L. (2016) Chromosomal-level assembly of the Asian seabass genome using long sequence reads and multi-layered scaffolding. *PLoS Genetics* **12**, e1005954.
- Voorrips R. (2002) MapChart: software for the graphical presentation of linkage maps and QTLs. *Journal of Heredity* **93**, 77-8.
- Vos P., Hogers R., Bleeker M., Reijans M., Lee T.v.d., Hornes M., Friters A., Pot J., Paleman J. & Kuiper M. (1995) AFLP: a new technique for DNA fingerprinting. *Nucleic Acids Research* **23**, 4407-14.
- Waldbieser G.C., Bosworth B.G., Nonneman D.J. & Wolters W.R. (2001) A microsatellite-based genetic linkage map for Channel catfish, *Ictalurus punctatus*. *Genetics* **158**, 727-34.
- Wall R., Ross R.P., Fitzgerald G.F. & Stanton C. (2010) Fatty acids from fish: the anti-inflammatory potential of long-chain omega-3 fatty acids. *Nutrition Reviews* **68**, 280.
- Wang B., Gan Z., Cai S., Wang Z., Yu D., Lin Z., Lu Y., Wu Z. & Jian J. (2016) Comprehensive identification and profiling of Nile tilapia (*Oreochromis niloticus*) microRNAs response to *Streptococcus agalactiae* infection through high-throughput sequencing. *Fish & Shellfish Immunology* **54**, 93-106.
- Wang C.M., Bai Z.Y., He X.P., Lin G., Xia J.H., Sun F., Lo L.C., Feng F., Zhu Z.Y. & Yue G.H. (2011) A high-resolution linkage map for comparative genome analysis and QTL fine mapping in Asian seabass, *Lates calcarifer*. *BMC Genomics* **12**.
- Wang L., Liu P., Huang S., Ye B., Chua E., Wan Z.Y. & Yue G.H. (2017) Genome-wide association study identifies loci associated with resistance to viral nervous necrosis disease in Asian seabass. *Marine Biotechnology* **19**, 255-65.
- Wang L., Wan Z.Y., Bai B., Huang S.Q., Chua E., Lee M., Pang H.Y., Wen Y.F., Liu P., Liu F., Sun F., Lin G., Ye B.Q. & Yue G.H. (2015a) Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Scientific Reports* **5**, 16358.
- Wang M. & Lu M. (2016) Tilapia polyculture: a global review. *Aquaculture Research* **47**, 2363-74.
- Wang S., Peatman E., Abernathy J., Waldbieser G., Lindquist E., Richardson P., Lucas S., Wang M., Li P., Thimmapuram J., Liu L., Vullaganti D., Kucuktas H., Murdock C., Small B.C., Wilson M., Liu H., Jiang Y., Lee Y., Chen F., Lu J., Wang W., Xu P., Somridhivej B., Baoprasertkul P., Quilang J., Sha Z., Bao B., Wang Y., Wang Q., Takano T., Nandi S., Liu S., Wong L., Kaltenboeck L., Quiniou S., Bengten E., Miller N., Trant J., Rokhsar D., Liu Z. & the Catfish Genome C. (2010) Assembly of 500,000 inter-specific catfish expressed sequence tags and large scale gene-associated marker development for whole genome association studies. *Genome Biology* **11**, R8-R.
- Wang W., Hu Y., Ma Y., Xu L., Guan J. & Kong J. (2015b) High-density genetic linkage mapping in Turbot (*Scophthalmus maximus* L.) based on SNP markers and major sex- and growth-related regions detection. *Plos One* **10**, e0120410.
- Wang Z., Gerstein M. & Snyder M. (2009) RNA-Seq: a revolutionary tool for transcriptomics. *Nature Reviews Genetics* **10**, 57-63.
- Watanabe T. (1982) Lipid nutrition in fish. *Comparative Biochemistry and Physiology Part B: Comparative Biochemistry* **73**, 3-15.
- Watanabe W., Ellingson L., Wicklund R. & Olla B. (1988) The effects of salinity on growth, food consumption and conversion in juvenile, monosex male Florida red tilapia. In: *The Second*

- International Symposium on Tilapia in Aquaculture. ICLARM Conference Proceedings*, pp. 515-23.
- Watanabe W.O., Kuo C.-M. & Huang M.-C. (1985) Salinity tolerance of Nile tilapia fry (*Oreochromis niloticus*), spawned and hatched at various salinities. *Aquaculture* **48**, 159-76.
- Watt K.I., Turner B.J., Hagg A., Zhang X., Davey J.R., Qian H., Beyer C., Winbanks C.E., Harvey K.F. & Gregorevic P. (2015) The Hippo pathway effector YAP is a critical regulator of skeletal muscle fibre size. *Nature Communications* **6**, 6048.
- Weatherley A.H., Gill H.S. & Lobo A.F. (1988) Recruitment and maximal diameter of axial muscle fibres in teleosts and their relationship to somatic growth and ultimate size. *Journal of Fish Biology* **33**, 851-9.
- Weber J.L. & May P.E. (1989) Abundant class of human DNA polymorphisms which can be typed using the polymerase chain reaction. *American Journal of Human Genetics* **44**, 388-96.
- Webster C.D. & Lim C. (2006) *Tilapia: biology, culture, and nutrition*. CRC Press.
- Wei L., Yang C., Tao W. & Wang D. (2016) Genome-wide identification and transcriptome-based expression profiling of the Sox gene family in the Nile tilapia (*Oreochromis niloticus*). *International Journal of Molecular Sciences* **17**, 270.
- Weichselbaum E., Coe S., Buttriss J. & Stanner S. (2013) Fish in the diet: A review. *Nutrition Bulletin* **38**, 128-77.
- Weller J.I. (2009) *Quantitative trait Loci analysis in animals (2nd Ed)*. CABI.
- Whatmore P., Nguyen N.H., Miller A., Lamont R., Powell D., D'Antignana T., Bubner E., Elizur A. & Knibb W. (2013) Genetic parameters for economically important traits in yellowtail kingfish *Seriola lalandi*. *Aquaculture* **400**, 77-84.
- Wilhelm B.T. & Landry J.-R. (2009) RNA-Seq—quantitative measurement of expression through massively parallel RNA-sequencing. *Methods* **48**, 249-57.
- Williams J.G., Kubelik A.R., Livak K.J., Rafalski J.A. & Tingey S.V. (1990) DNA polymorphisms amplified by arbitrary primers are useful as genetic markers. *Nucleic Acids Research* **18**, 6531-5.
- Wilson K., Li Y., Whan V., Lehnert S., Byrne K., Moore S., Pongsomboon S., Tassanakajon A., Rosenberg G., Ballment E., Fayazi Z., Swan J., Kenway M. & Benzie J. (2002) Genetic mapping of the black tiger shrimp *Penaeus monodon* with amplified fragment length polymorphism. *Aquaculture* **204**, 297-309.
- Witczak C.A., Sharoff C.G. & Goodyear L.J. (2008) AMP-activated protein kinase in skeletal muscle: From structure and localization to its role as a master regulator of cellular metabolism. *Cellular and Molecular Life Sciences* **65**, 3737-55.
- Wolff C., Roy S. & Ingham P.W. (2003) Multiple muscle cell identities induced by distinct levels and timing of hedgehog activity in the Zebrafish embryo. *Current Biology* **13**, 1169-81.
- Wringe B.F., Devlin R.H., Ferguson M.M., Moghadam H.K., Sakhrani D. & Danzmann R.G. (2010) Growth-related quantitative trait loci in domestic and wild rainbow trout (*Oncorhynchus mykiss*). *BMC Genetics* **11**, 63.
- Wu N., Song Y.-L., Wang B., Zhang X.-Y., Zhang X.-J., Wang Y.-L., Cheng Y.-Y., Chen D.-D., Xia X.-Q., Lu Y.-S. & Zhang Y.-A. (2016) Fish gut-liver immunity during homeostasis or inflammation revealed by integrative transcriptome and proteome studies. *Scientific Reports* **6**, 36048.
- Wu R., Ma C. & Casella G. (2007) *Statistical genetics of quantitative traits: linkage, maps and QTL*. Springer Science & Business Media.
- Xia J.H., Bai Z., Meng Z., Zhang Y., Wang L., Liu F., Jing W., Yi Wan Z., Li J., Lin H. & Hua Yue G. (2015) Signatures of selection in tilapia revealed by whole genome resequencing. *Scientific Reports* **5**, 14168.
- Xia J.H., Lin G., He X., Yunping B., Liu P., Liu F., Sun F., Tu R. & Yue G.H. (2014a) Mapping quantitative trait loci for omega-3 fatty acids in Asian seabass. *Marine Biotechnology* **16**, 1-9.
- Xia J.H., Liu F., Zhu Z.Y., Fu J., Feng J., Li J. & Yue G.H. (2010) A consensus linkage map of the grass carp (*Ctenopharyngodon idella*) based on microsatellites and SNPs. *BMC Genomics* **11**, 135.
- Xia J.H., Liu P., Liu F., Lin G., Sun F., Tu R. & Yue G.H. (2013) Analysis of Stress-Responsive Transcriptome in the Intestine of Asian Seabass (*Lates calcarifer*) using RNA-Seq. *DNA Research* **20**, 449-60.
- Xia J.H., Wan Z.Y., Ng Z.L., Wang L., Fu G.H., Lin G., Liu F. & Yue G.H. (2014b) Genome-wide discovery and in silico mapping of gene-associated SNPs in Nile tilapia. *Aquaculture* **432**, 67-73.
- Xiao S., Wang P., Zhang Y., Fang L., Liu Y., Li J.-T. & Wang Z.-Y. (2015) Gene map of large yellow croaker (*Larimichthys crocea*) provides insights into teleost genome evolution and conserved regions associated with growth. *Scientific Reports* **5**, 18661.
- Xiong Y., Huang J., Li X., Zhou L., Dong F., Ye H. & Gan L. (2014) Deep sequencing of the tilapia (*Oreochromis niloticus*) liver transcriptome response to dietary protein to starch ratio. *Aquaculture* **433**, 299-306.
- Xu J., Zhao Z., Zhang X., Zheng X., Li J., Jiang Y., Kuang Y., Zhang Y., Feng J., Li C., Yu J., Li Q., Zhu Y., Liu Y., Xu P. & Sun X. (2014a) Development and evaluation of the first high-throughput SNP array for common carp (*Cyprinus carpio*). *BMC Genomics* **15**, 307.
- Xu P., Zhang X., Wang X., Li J., Liu G., Kuang Y., Xu J., Zheng X., Ren L., Wang G., Zhang Y., Huo L., Zhao Z., Cao D., Lu C., Li C., Zhou Y., Liu Z., Fan Z., Shan G., Li X., Wu S., Song L., Hou G., Jiang Y., Jeney Z., Yu D., Wang L., Shao C., Song L., Sun J., Ji P., Wang J., Li Q., Xu L., Sun F.,

- Feng J., Wang C., Wang S., Wang B., Li Y., Zhu Y., Xue W., Zhao L., Wang J., Gu Y., Lv W., Wu K., Xiao J., Wu J., Zhang Z., Yu J. & Sun X. (2014b) Genome sequence and genetic diversity of the common carp, *Cyprinus carpio*. *Nature Genetics* **46**, 1212-9.
- Xu Z., Gan L., Li T., Xu C., Chen K., Wang X., Qin J.G., Chen L. & Li E. (2015) Transcriptome profiling and molecular pathway analysis of genes in association with salinity adaptation in Nile tilapia *Oreochromis niloticus*. *Plos One* **10**, e0136506.
- Yamamoto T., Shima T., Furuita H. & Suzuki N. (2002) Influence of dietary fat level and whole-body adiposity on voluntary energy intake by juvenile rainbow trout *Oncorhynchus mykiss* (Walbaum) under self-feeding conditions. *Aquaculture Research* **33**, 715-23.
- Yan B., Guo J.-T., Zhao L.-H. & Zhao J.-L. (2012a) microRNA expression signature in skeletal muscle of Nile tilapia. *Aquaculture* **364-365**, 240-6.
- Yan B., Guo J.-T., Zhao L.-H. & Zhao J.-L. (2012b) MiR-30c: A novel regulator of salt tolerance in tilapia. *Biochemical and Biophysical Research Communications* **425**, 315-20.
- Yan B., Guo J.-T., Zhu C.-d., Zhao L.-H. & Zhao J.-L. (2013a) miR-203b: a novel regulator of MyoD expression in tilapia skeletal muscle. *The Journal of Experimental Biology* **216**, 447-51.
- Yan B., Liu B., Zhu C.-D., Li K.-L., Yue L.-J., Zhao J.-L., Gong X.-L. & Wang C.-H. (2013b) microRNA regulation of skin pigmentation in fish. *Journal of Cell Science* **126**, 3401-8.
- Yan B., Zhu C.-D., Guo J.-T., Zhao L.-H. & Zhao J.-L. (2013c) miR-206 regulates the growth of the teleost tilapia (*Oreochromis niloticus*) through the modulation of IGF-1 gene expression. *The Journal of Experimental Biology* **216**, 1265-9.
- Yan X., Ding L., Li Y., Zhang X., Liang Y., Sun X. & Teng C.-B. (2012c) Identification and profiling of microRNAs from skeletal muscle of the common carp. *Plos One* **7**, e30925.
- Yáñez J.M., Naswa S., López M.E., Bassini L., Correa K., Gilbey J., Bernatchez L., Norris A., Neira R., Lhorente J.P., Schnable P.S., Newman S., Mileham A., Deeb N., Di Genova A. & Maass A. (2016) Genomewide single nucleotide polymorphism discovery in Atlantic salmon (*Salmo salar*): validation in wild and farmed American and European populations. *Molecular Ecology Resources* **16**, 1002-11.
- Ye J., Coulouris G., Zaretskaya I., Cutcutache I., Rozen S. & Madden T.L. (2012) Primer-BLAST: A tool to design target-specific primers for polymerase chain reaction. *BMC Bioinformatics* **13**, 134.
- You X., Shu L., Li S., Chen J., Luo J., Lu J., Mu Q., Bai J., Xia Q., Chen Q., Cai Y., Zhang H., Chen G., Lin H., Zhang Y. & Shi Q. (2013) Construction of high-density genetic linkage maps for orange-spotted grouper *Epinephelus coioides* using multiplexed shotgun genotyping. *BMC Genetics* **14**, 113.
- Young K. (2009) Omega-6 (n-6) and omega-3 (n-3) fatty acids in tilapia and human health: a review. *International Journal of Food Sciences and Nutrition* **60**, 203-11.
- Young N.D. (1996) QTL mapping and quantitative disease resistance in plants. *Annual Review of Phytopathology* **34**, 479-501.
- Yu F.-X. & Guan K.-L. (2013) The Hippo pathway: regulators and regulations. *Genes & Development* **27**, 355-71.
- Yu Y., Zhang X., Yuan J., Li F., Chen X., Zhao Y., Huang L., Zheng H. & Xiang J. (2015) Genome survey and high-density genetic map construction provide genomic and genetic resources for the Pacific white shrimp *Litopenaeus vannamei*. *Scientific Reports* **5**, 15612.
- Yu Z. & Guo X. (2003) Genetic linkage map of the eastern oyster *Crassostrea virginica* gmelin. *The Biological Bulletin* **204**, 327-38.
- Yue G., Lin H. & Li J. (2016) Tilapia is the fish for next-generation aquaculture. *International Journal Marine Science Ocean Technology* **3**, 11-3.
- Yue G.H. (2014) Recent advances of genome mapping and marker-assisted selection in aquaculture. *Fish and Fisheries* **15**, 376-96.
- Yue G.H. & Orban L. (2005) A simple and affordable method for high-throughput DNA extraction from animal tissues for polymerase chain reaction. *Electrophoresis* **26**, 3081-3.
- Yue G.H. & Wang L. (2017) Current status of genome sequencing and its applications in aquaculture. *Aquaculture* **468**, 337-47.
- Zeng Z.B. (1993) Theoretical basis for separation of multiple linked gene effects in mapping quantitative trait loci. *Proceedings of the National Academy of Sciences* **90**, 10972-6.
- Zeng Z.B. (1994) Precision mapping of quantitative trait loci. *Genetics* **136**, 1457-68.
- Zhang G., Yin S., Mao J., Liang F., Zhao C., Li P., Zhou G., Chen S. & Tang Z. (2016) Integrated analysis of mRNA-seq and miRNA-seq in the liver of *Pelteobagrus vachelli* in response to hypoxia. *Scientific Reports* **6**, 22907.
- Zhang L., Yang C., Zhang Y., Li L., Zhang X., Zhang Q. & Xiang J. (2007) A genetic linkage map of Pacific white shrimp (*Litopenaeus vannamei*): sex-linked microsatellite markers and high recombination rates. *Genetica* **131**, 37-49.
- Zhang Q., Xie D., Wang S., You C., Monroig Ó., Tocher D.R. & Li Y. (2014) miR-17 is involved in the regulation of LC-PUFA biosynthesis in vertebrates: Effects on liver expression of a fatty acyl desaturase in the marine teleost *Siganus canaliculatus*. *Biochimica et Biophysica Acta (BBA) - Molecular and Cell Biology of Lipids* **1841**, 934-43.

- Zhang R., Zhang L.-l., Ye X., Tian Y.-y., Sun C.-f., Lu M.-x. & Bai J.-j. (2013) Transcriptome profiling and digital gene expression analysis of Nile tilapia (*Oreochromis niloticus*) infected by *Streptococcus agalactiae*. *Molecular Biology Reports* **40**, 5657-68.
- Zhao B., Tumaneng K. & Guan K.-L. (2011) The Hippo pathway in organ size control, tissue regeneration and stem cell self-renewal. *Nature cell biology* **13**, 877-83.
- Zhao Y., Wang J., Thammaratsuntorn J., Wu J., Wei J., Wang Y., Xu J. & Zhao J. (2015) Comparative transcriptome analysis of Nile tilapia (*Oreochromis niloticus*) in response to alkalinity stress. *Genetics and Molecular Research* **14**, 17916-26.
- Zhao Y., Wu J.-W., Wang Y. & Zhao J.-L. (2016) Role of miR-21 in alkalinity stress tolerance in tilapia. *Biochemical and Biophysical Research Communications* **471**, 26-33.
- Zhao Y., Zhu C.-D., Yan B., Zhao J.-L. & Wang Z.-H. (2014) miRNA-directed regulation of VEGF in tilapia under hypoxia condition. *Biochemical and Biophysical Research Communications* **454**, 183-8.
- Zhou T., Liu S., Geng X., Jin Y., Jiang C., Bao L., Yao J., Zhang Y., Zhang J., Sun L., Wang X., Li N., Tan S. & Liu Z. (2017) GWAS analysis of QTL for enteric septicemia of catfish and their involved genes suggest evolutionary conservation of a molecular mechanism of disease resistance. *Molecular Genetics and Genomics* **292**, 231-42.
- Zhu J., Li C., Ao Q., Tan Y., Luo Y., Guo Y., Lan G., Jiang H. & Gan X. (2015) Transcriptomic profiling revealed the signatures of acute immune response in tilapia (*Oreochromis niloticus*) following *Streptococcus iniae* challenge. *Fish & Shellfish Immunology* **46**, 346-53.
- Zhu W., Wang L., Dong Z., Chen X., Song F., Liu N., Yang H. & Fu J. (2016) Comparative transcriptome analysis identifies candidate genes related to skin color differentiation in Red tilapia. **6**, 31347.
- Zhu Z.Y., Lo L.C., Lin G., Xu Y.X. & Yue G.H. (2005) Isolation and characterization of polymorphic microsatellites from red coral grouper (*Plectropomus maculatus*). *Molecular Ecology Notes* **5**, 579-81.
- Zong G., Li Y., Wanders A.J., Alsema M., Zock P.L., Willett W.C., Hu F.B. & Sun Q. (2016) Intake of individual saturated fatty acids and risk of coronary heart disease in US men and women: two prospective longitudinal cohort studies. *British Medical Journal* **355**, i5796.
- Zou Z., Li D., Zhu J., Han J., Xiao W. & Yang H. (2015) Genetic variation among four bred populations of two tilapia strains, based on mitochondrial D-loop sequences. *Mitochondrial DNA* **26**, 426-30.

Appendix 1. Selected microsatellite markers for reconstruction of a linkage map in salt-tolerant tilapia

Linkage	Marker				Product length	Annealing
group	name	Primer sequence (forward)	Primer sequence (reverse)	Motif	(bp)	temperature
1	GM633	GACCCAGGACTCATGTGCTT	GTGTCCCAAGAAAACCAGGA	(CA)27	178	55
1	OMO165	AAAGCGAGGATCAGCCAGAGTTAC	TGACGGGGTGAACAAAATAAAAGA	(CA)10	287	60
1	OMO188	ACATATTCGCCTTGTTCTCACCAG	GAAGTGTGCGTTCAAAGTATGATT	(CA)16	163	55
1	OMO287	AACCTGGTTAGTCTCTGGCTGCTC	ATGGCATGGTTTGCAACATAATCC	(GT)17	280	60
1	OMO293	CTATCTGCCCCGGTTTCTTTGT	ATGCCGCATTTTTAGTTTCCATA	(GA)23	435	60
1	OMO432	AGTAGTTTTTGTTCATTAGAGACG	TAGTCACACTGATAATTTTTCTTG	(ATAG)15	224	50
2	OMO048	TCCCCACACCTCCTGCTCCTC	AGCCCAAGCCAGTCCCATAGC	(CA)18	336	60
2	OMO160	AGGATTTCTTGAAAGTGTTTTT	ACTCTACGTGACCTCTGACAATAG	(GT)10	210	55
2	OMO177	AGTGATGACCGGCCAGAAAGAGA	CAGGGATGGATAAACGTGACAATG	(CA)11	345	55
2	OMO248	AAAGACACAAAGAGAAACTAATCA	GGATGAATATTTAAAATCAGTCAG	(TCA)9	381	55
2	OMO272	CAGGGCTCATGGTAACAGTTCATA	TAGTACGAAGCGCTAGCAGTCTCT	(GT)23	454	60
2	OMO317	ACATCTACTGGTATCTGGGTTCAA	GTTTAATTCTGGGAGGGCACAC	(CA)32	443	60
3	GM139	TTTGAGTAACCACCCTAACAC	GTGGGATCTACCAAGAAGAG	(GT)17	255	55
3	GM150	AGGTGATTGGCTTAGATGAT	GTCTCAGTTTGTTTGGCTTA	(CA)30	177	55
3	GM385	GTAAAGCAAACTCAGAGGGT	AATGATCCCGGCTGTT	(CA)10	282	55
3	OMO186	CAAATGAGCGAAGAAAACAATAAA	TCCCTCTCCCATCTGAGTCAAGTA	(GT)10	333	55
3	OMO241	AGAGTACGACCCTGCACACTTCAC	TGATCAACAATCTGGCTTATTCTG	(CA)23	219	50

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
3	OMO274	TCAGCTGGATCTGCTCAGTGTGAC	ACAGAACGCAGTGCTTTCATTTTG	(CA)9	401	60
3	OMO388	GGGATTTGTTTTCTCCTTAGCA	CAGCCATTGTTGTGTATTTGTGAC	(GT)7	143	55
4	GM509	AGCCCTGCACAGTTTAGCAA	CAACAGCAAAGCCAGGAGAG	(CA)23	181	55
4	OMO075	CGGCTCGGCCTCCTCATCT	TAGGCACAAAGTGGGTGGAAGGT	(CA)13	464	55
4	OMO092	GCACGCTGAAACTTACACAGTGA	GAGGCTTAGTTGGGCTGATTGATT	(CT)12	185	60
4	OMO224	TTCCCGGAGTACAGAATGTTAGAC	GAGGCTGCATTAGCTAGTGGATAC	(GT)13	260	55
4	OMO226	ACGAGCCAACCTGCGAAGAGAT	GCTGCTCAAGCCAAAAATGTCAAA	(GT)22	451	55
4	OMO299	GTTGGACAACATCGCCTACATCAT	CACGCAATTAACGTCCCCTGAA	(GT)10	365	60
4	OMO354	AAGTATAAACCCCGTGCGTGAG	ACAGGGGAGGTGGAGGAGGAA	(CA)14	342	60
5	GM006	AGCCAACAACCTATTCTGCTGT	TCGGGTTGTGCTCCTTAG	(CA)16	93	55
5	GM017	GATACCTGTCCATACCTCCTC	CCCTCTGTTTCCATCTCA	(CA)26	160	55
5	OMO006	TTGGCCTTTTCTCTGGGGTTATT	TGGCCCGTTAACTCTGAAAACAT	(CA)10	435	55
5	OMO024	AGTTTCTGCTGCTATTCATTAC	GGCTCAAGATACCCATACAA	(CA)27	403	60
5	OMO155	AGATCTCCCCGAAACCCTGATTAT	CAAACGCTACGCTGTGCTGCTA	(CA)12	206	60
5	OMO192	TTTCCGCACATTAATCTCAATACA	ATCTTTCATCTGCTCACCGTCTCT	(AT)9	399	60
6	GM440	CTGCACTTTTACTGAGGG	TGGGAGATTAACAGAATAAC	(CA)21	264	55
6	OMO059	TCCACGATGGGATGAACCTTGAG	TGTTGCCGTGTGACTGTTGCTCT	(CA)10	179	60
6	OMO067	GTTTCTCGTCCCAAAGTCTCACAG	TCCAGATTCGCTTTAGTTTGTTC	(GT)8	179	60
6	OMO265	CAACCACTGCAACCACGTAAAGAG	CCCAGGCCAGCTGACAGAAA	(GT)8	110	60

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
6	OMO303	CTCACCCGAGCAGGCAGGAG	GCTGATGTTCTGGGGGATAATGG	(CAG)7(CAA)10	181	55
6	OMO306	AATGCGAAGGCGAGTGGAGTG	AGCAGGTGATTAGGAGAGGCTGAT	(GT)10	182	60
6	OMO374	GTAAGCGCAATGATATGAATGAAT	TTGCCAAGATATTAGAACCAGATG	(CA)12	409	60
7	OMO045	ATCTCCCAAAGACATGCAAAAGAA	ATAAGCCTCAGTACCCCCACCACT	(AT)8	417	60
7	OMO065	AACGGTTTAACAGTCACGCACATC	TGCTTACTCGCCATTGACATTCTT	(GT)9	240	60
7	OMO171	AGCCTCGGGAAGAGTGAAGTCAAC	TTTTTGCTAACCCATCTCCTGTGA	(CT)9	153	55
7	OMO233	TGTTCAGACTGCAGCTTCATTAGA	CAGCCTGCTGGTCTGGAAATA	(GT)19	336	60
7	OMO313	GGGGGTGCAATAAAGGCTGAG	CTGCAAGGCACAACTATCAATAA	(TA)8	445	60
7	OMO361	TGACAGCGAGCCAGAATGGAAGTA	AAAAGTGAAAGGGGCACAGTGAGG	(CTT)17	450	55
8	GM104	ACTAACCCTGCTCTGTGCTT	GAACCCAGCGATGTCCC	(GT)15	273	55
8	GM559	ATTTTCCACCGAGCTCACCT	GCACCATTTTAAACCAGTGC	(GT)24	134	55
8	OMO158	CTTGCAGACGGTTCATTTACGAG	TCCGGTTAATCGCTCTCACTCTCA	(GT)18	425	55
8	OMO238	ATATTACGTCCAAACATCCAGAGC	GAGCCAAAGGCAGAAGTAAACAGT	(CT)11	208	55
8	OMO255	GAGTTTTGAGCCAGCATGATTTCA	AAAAGGGGACAGGTCACTGGATTC	(AT)7	356	60
8	OMO277	TGAAGCAGTGTGAGCAGAGG	TTTTTCCTTTTAGGTGTTTGAG	(GAA)9	133	55
8	OMO312	ATAGTTTGGCAGGTCATTTTCAGA	GGGGTAGTTTGTGTGCTTTTT	(GT)30	345	55
9	GM032	TTTCGACAAGAAGGCGTTGA	CGCAGAGATTTTGGCTCATT	(CA)15	244	55
9	GM171	GAGTTAGGGCAGAGGTTATGT	GCCATCACCTCTTGCTT	(CA)22	280	55
9	GM184	GGAGAACAGGGCTTATTG	CACTTCCACCTTTACTTTGA	(CA)22	209	55

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
9	GM253	CAGAGTGATGCGTCTTCC	GCAAATACAGCACCATTCTA	(CA)24	218	55
9	GM566	TCTTCCGTTTGACTCACCAA	CAGACACGCAGAGATGTGGA	(CA)38	178	55
9	OMO167	TCACGGACACTGAACAAAGAGGA	TGGCCATGGAGATTTAAAGAACAC	(CA)23	393	60
9	OMO337	TAGGAGAGGCATAGGTTGTCAAAT	CAAGAGTCTAGGAGGGAATCAAAA	(GTTT)7	151	55
10	GM135	ATGGAAGGTCGCCCTAATAAG	GGTCCAGACACGGTCGTC	(CA)27	262	55
10	OMO220	ACCGCCCCTTCCTCTACCTGA	TGTGTTTGTGTATCCGCCGAGAT	(CA)13	408	55
10	OMO222	AAGCGATAACGGAGATTCTTCTGA	TAAAGGGTGCCGGTAAATGTAAAT	(GT)9	178	55
10	OMO250	TCCGGCTGCTCCAGTGTGC	GGCTGACGGAAAAATGGCTCAT	(GT)15	428	60
10	OMO300	TTTGCACCCTGTACTCTCACTGTA	ATATCGGGTAACATTCGCTCCACT	(CA)17	399	60
10	OMO320	TTTTTGGCAAGCTGGTGTAGGA	TGGACAAAGTTATGGGACCTCTGA	(GTA)8	367	60
11	GM131	AAAATGGCATGAAAGAACTAA	GCCATTGATTTGTTGACACT	(CA)21	227	55
11	GM333	CTGCGTGAATTAGCAAAAATA	AAAATAAGCAAAGTGCCATA	(CA)23	176	55
11	OMO205	GAGAGGGAGGGATGCTTGTT	TTTGAAATTCCCACATTACATCTT	(GT)11	269	55
11	OMO261	GCACCGATTAGTTTGCTTCATT	CAACTGCTTTGGGGTTTATTTTAC	(CT)24	107	60
11	OMO295	CAAAGCGATTGATCGGAGTCTGG	CTCCCTCTCAAAACCCGATGTTAT	(ATG)7(ATT)9	434	60
11	OMO380	CTCGTCCTGCGTTGCCTTTTT	CAACAGCCTCCTGACCTTTTCCTA	(TTA)7	453	60
12	GM284	CGCAGCTTTGATTACATACCT	AGCCGGTTTCCAGTGAGAA	(CA)16	177	55
12	GM369	TTTAACGAGTCTGGCATCTGA	TGCATTCATCCCATCACC	(CA)25	208	55
12	OMO166	CACCTCATCGGGAGACAGACAAG	TGATGGCCACTAACTCGTGAAAAA	(CA)15	307	60
12	OMO285	GACAATCACTCCTCCAGCCTTACC	CAGACGCTCAACTCACAGGAAAAA	(CA)10	285	60

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
12	OMO346	GACGTGAGCGGAGCTTTGTTTG	CTTCTGATGCGGTGAGGTGTTTTT	(GT)13	451	60
12	OMO350	GCCAACTTTCATGTACATACCAAT	ATATCATTCCCAAGTGCTCTCATT	(AT)8(GT)6	327	60
12	OMO402	GGGTGGAGGTTTTAACAGTTGATG	TTGGCTCATTTTGTACCGTGATG	(GT)9	369	60
12	OMO412	TTTTTGTGGATTTTCGGTGAGG	GCAGCGTACCAGAGTTCAGG	(GT)8	278	55
13	OMO032	GGTAAGCGGTGCAAAGTTCA	GTTGGCTAACCTCTGTTCCTCC	(GAAAA)10	313	55
13	OMO116	TGAAAATCAGCAGCACTTCGTT	TTGGGGTATTTTATGTTGGCATCT	(CA)9	365	60
13	OMO184	GGGGGCATGATTTACTTCTGAGGT	AAGGCTGGAGCACAGGGACTTT	(CA)12	182	60
13	OMO232	CATGAGCAAACAACAAACAGTGAA	GAGGCAAAGGAAGCAAGGTGAAG	(GAAA)9	470	60
13	OMO369	GAGTCCCAGGACCTCTCAACC	TCATAGTGCAGCCCAATAGTGTCA	(CA)9	466	60
13	OMO391	AGACATCTGTACGCTCTTTACGAA	AGTGCTAGAGGGAAGGGGCTGTA	(GAT)9	290	60
13	OMO424	GTCGGCTGATGGGTGTAGTGGAG	AAAGGCCGTGGGGAGGTGAG	(CA)13	146	60
14	GM261	GCTGTAGTTACCACGAGGTGA	GTTTGGCAGCGTGTCTCAG	(GT)33	131	55
14	GM276	TCAAGTTGCGTCTCTGTCACC	CGCAGGAGGCTTTACCACA	(GT)16	247	55
14	OMO127	ATTTGGGTCATGAATTGGGTTTG	TGGACTGTTGCCCTTTAAGTGTTT	(GT)15	359	60
14	OMO182	CTCTCACTTTCACTGGGAAGGATA	AGTTCTTAGGATGCGGTCTGTCTG	(CT)20	190	55
14	OMO262	GCGTAGCAGCGGTGGTGTATTTA	CACCCCTGCAGTCATCCTCTACTT	(CA)10	161	60
14	OMO372	ATGGGATATTTTTGTCAATGAAGA	CAAGTGGCGTTAAGGTGTCAGTAG	(AT)9	304	60
15	OMO051	CAGCGGAGGAGGAGTCACAGG	AGTGGGAACAAGGGGATTAGACAT	(CT)18	278	60
15	OMO069	GGGCCGCCTCCTCCTTACAT	GCGGGTGGGAGGGTTGAACTA	(CA)16	194	60
15	OMO074	CAACTGTTTAAGGGCTTTGATGTT	AAGGCGACTTTTCTGAGATTGATA	(CA)10	429	55

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
15	OMO099	TTCCAGAATGGCAACAATAATGTA	AATAAATGTGTTCGCCACTGTAAT	(GT)15	234	60
15	OMO379	ATCCTCGCTGTCATCATCTCACTC	GAAACCAGCTTCACTGATAACAAA	(GT)13	395	60
15	OMO409	CTTGACTATGTAACGCATTCACCTT	ATGATAGTAACGGCTGTTCCCTGTG	(CA)8	334	55
15	OMO416	TCGGGAGGTGGAAGAACGATACA	CCACCCACCAAGTGAAGTGTGC	(GT)9	455	55
16	GM581	CAAAAGCAGCCACAAGTCAA	ACCTATGTGGGCTGAATGCT	(CA)34	150	55
16	GM609	GCACGGTCCTTACTGGTTGT	TCCCAGAGTGCCTTTGTTCT	(GT)22	106	55
16	OMO130	CTTGCTCTAGGATGATCTTCTGTG	TGAATGCAGCTGTGGTACAACCTCC	(GT)11	140	55
16	OMO366	TGCTATGCTTGCTCAACCTTAGAT	ATGGCACTTGACAGAAACCTTTTA	(GT)20	271	60
16	OMO378	TTAATATTGGCACAAAAGTCTGGA	AGTCAACTGATGATGCACAAAAAG	(TA)11	337	60
16	OMO397	ACGCGTGTTTGAGATATTTAGATT	GAACAAACAAGGGGAGTGG	(GATT)7	222	60
16	OMO407	TGTGCAGCTTTCTTGGAGTTTCTA	TGGATTACTTAACGCAGCCTAACA	(GT)11	336	60
17	GM342	AAAGCCAAAACCAGCACA	TCAATCCTCTGTCACCCATA	(GT)13	264	55
17	GM655	AACCCTGTTGGCTGTCTCAC	CCGACAGTCAGCTCAGATCA	(CA)32	101	55
17	OMO330	TCAGAAGCAGCTTGAAAAATCA	TTTATTCCATGGTGAAAACATTCC	(GT)14	424	60
17	OMO382	CCATCCGCAGCTTCATTTTCATAC	AGCCAGAGCTGGAGGGACGATA	(CT)9	417	60
18	OMO080	CCAACCGACACATGTAAC TCAAAA	AAACGGGTAGCTGCTGCTCTATGA	(GT)13	223	60
18	OMO131	TCCCTTAAAATATTACTGCTCCTC	AGTCCCATTGCTTTACTACAGAC	(CA)13	139	55
18	OMO236	GGACCATGCCATCCATCT	CGCCACAGCGATCTGACTA	(GT)14	127	60
18	OMO327	TTATTTTCAAACCCAAACTCAGC	CATGGGGTGTTATCAGATGTGC	(GT)12	272	55
18	OMO341	TGGAGCTCTACTTTGCCCTACTA	ACGCTATAGATGGACCCTGGATTT	(CA)24	369	60

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
18	OMO426	ATGCGTGGTTATTAGGTGTGGTAT	TAATAGGATCGGTGACTTCAAACA	(AAGG)8	143	55
19	GM112	ATCCCGCTAACAGACTCA	CCGATATTCATCACGTTTAA	(CA)16	204	55
19	GM175	GGGGCTGTAAGTGTCTGTATG	GCAAAGTACCAAATGACC	(CA)27	192	55
19	GM634	CCCGACATTAAACTTTTACGA	CTGAAACATGACTGCAGGAG	(GT)26	193	55
19	OMO081	ACAGCCCTGCTCCATTCATCTTTA	GCTCAGCCAGAGGGAAAAACACTA	(CA)8	457	55
19	OMO210	CCCCCAACACGCCTCAGAAG	GGTGCCAAAATAAGACCCAGTGC	(CA)9	150	55
19	OMO281	TCTTCACCGCTTGTATCTGTCACG	AAGGCTTGAAGCACATGAGTGGAT	(GT)25	293	60
19	OMO310	CAGTTAACTGGTAATTGCCTTTTG	ATTCTGCGTTGGTTTCTACTTTTT	(GA)13	151	60
20	GM122	AAAGGTCACAAAGGCAGCAAA	AGCCCCGCCACAATAG	(CA)11	172	55
20	OMO097	TTTTTCCCCCACCGTTCTTCTCT	GCTGCTCTAACGCTGAACACAACC	(CA)13	224	60
20	OMO129	TTGGCAGGCTAAGTACTATTTTCAT	GAGCGAATGGTTGTCTGTCTCT	(CCAT)9	267	55
20	OMO253	GGGTGGCGATCACATACTTTTCTA	CTCAAACAGTATGGTGGTGGATGG	(TA)8	243	55
20	OMO288	GTCCGGGCCCCTGACTCTAAC	GTCGTGCCAATGCTGTCACACTT	(GT)8	142	60
20	OMO302	TTTAGAGTCCTGGCTGGGTGTAAT	TGCTGTGTCACGGGACTCTTTTAG	(CT)16	318	55
20	OMO360	TGAATCCGGGCTAATTACCTGTA	GTATTATTGGCCAACAACGCACAC	(CA)23	374	60
21	OMO126	GGTCGTAGCGATTTAGTCCAAACT	TTTTATGCGGCATGACAGATTTTA	(CT)15	173	60
21	OMO209	GCGGCACTTCTTTACATCTTACGA	AGGCGCAAAGCAGCATCAAA	(CA)27	247	55
21	OMO247	CTTGAAGGGAAGCGCTGTAAATAC	GCAGGAAGACGGCTGGGTGTA	(CA)11	432	55
21	OMO301	AGCCATTCTCATTGATTGTCGTG	GGGGAGGGAGGGGGTTAGC	(CA)10	193	55
21	OMO367	CAGGCCAGAAAACAGCAAAACC	TGACCCTCATGTTCAGTTCAAAGA	(AAT)10(CA)11	346	60

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Motif	Product length (bp)	Annealing temperature
21	OMO375	TGGCTAAATTTAGAGCTCACTGTT	GAAGGCAGGTAATCTAGCAAAAAG	(GT)11	367	60
21	OMO392	CTGGCTTAACTTCTCTACTGGACA	TCTACTCAAACTGGCAACAAAAC	(GAATA)7	444	60
21	OMO395	AGATGCCTGAGCTGGAAAAGAGAT	TAAGCACCTGGATGGCATTATCAC	(CA)21	334	60
22	OMO038	ATACGCAGGAAATCTTCAAACACC	TTGCCCACTAGTGGTAAGAATCAA	(CAA)7	351	55
22	OMO049	AGCCAGCGGAGACCAACAAAAT	TCTCATGGGAGGCTCACTCACTCT	(GGAA)9	224	60
22	OMO106	CCGTTAAAGGAGCAGTCCAC	TTAGTGTTTCAATGCACAAATGAT	(CA)15	350	60
22	OMO107	AGCCCAAAACATCCCCTCTGTAA	CATGGGGTTTCTGTGACTCTTCCT	(CA)10	432	60
22	OMO244	TTTGCGTGTTCCCATCGTTTCTC	CTCCACATGCCGCCACAGC	(CT)10	434	55
22	OMO433	ATGTAGCGCTATCCACCTTCCATT	AAAAAGAGGCAGGCGAGTTCATAC	(CA)18	284	60

Appendix 2. Fine-mapping microsatellite marker pairs for male linkage group

Linkage group	Marker name	Primer sequence (forward)	Primer sequence (reverse)	Product length (bp)	Annealing temperature (°C)
2	2RM22.3b	CTTTGACACAGTAGCGTATGATGC	TCGGAAATTTGACTGTGACCTACT	368	55
	2RM22.3a	CCTCCATAATGTAATAAAGTCCTC	AAAGTTACGCAGTAGTAGTGTGA	362	55
	2RM20.6	CATGACTGGGCCAGCTCCATCCTA	GAGGCGGTGATTTGGGGGTAATTG	243	60
	2RM20.4	GGAAAGAAATGAAGGAAAACAAGA	GCCACCAGCATCACTCTCAC	338	60
	2RM19.0	TCATCAGTAAAAGGGCTAATGTGT	CTCTAAGCTGGATAGTCCCTTTCT	386	55
	2RM17.6b	CGCCGTGGGAAGCTGGTAGGA	ACGGTGGTGAAGCGAAATGGAGTG	401	60
18	18RM9.7b	GTGCCAGCGCTGCCACTCT	CTGTCAGAGGCGTAGAGGACGTTC	285	60
	18RM9.7a	AAAAACGGGGCCATGAGCATA	GGGCTGATGTAAGGCAGTCACAAG	198	60
	18RM21.6	TGTCCGCCCTACAGGCTACATTC	AACAAGGAGTCGGGAGAAGGAACA	264	55
	18RM20.2	AGCCAATTTAGCATCCCTCAGGA	GATCACGTGGGCTGAAAGCGTATT	408	55
	18RM18.8	CCAATGCGCTCTTCCACTTCTC	CAATTTGAGGCCGTTTGTATGGA	359	55
	18RM17.6b	GCCAATATTGCAGCAGGGTATCC	GAAAAAGAGCGGAAAACGCACAG	357	60
	18RM17.6a	GGGAATCTTCAGCGGCTGCTC	CCAGGGGTGACTGTGGGTCTCT	234	60
	18RM16.04	AATATCGGGGATCCTCTCACCATT	GATGAAACAGCAAGGGAGGAGAGT	390	60
	18RM16	GGTGTTTGGTTCCCCCTGAGA	TCAACCTCTACATCGGCATCTGG	330	55
	18RM13.8	GTTGAGCGTGCATGTGTCTATGGA	GTGCTTGGCCACAGTGTTCCTTA	216	55
	18RM10.4	GCAGGGGTCAAAAAGCAATTAAGA	ACCAGCAGCAGCTTCATAACCATC	331	60
	18RM10.3	GAAAATGGTCAAGCGGAAGG	AGGGCCGTGTCTGACTGCT	284	60

Appendix 3. List of primers used for real-time qPCR analyses of gene transcript analysis for the verification of differentially expressed genes in the muscle and liver of salt tolerant tilapia.

List of primers used for real-time qPCR analyses of gene transcript analysis in the muscle.		
Sequence (5'→3')		
Primer Name	Forward	Reverse
NTsmpxv2_1	TTGTTTCATCGACTGGTCAGCGG	TGGACGCAGAGCTCCCATCG
NTmyoz_1	ACCCCTGAGCCTCCGTGTCT	ACGTAACCGCTGGTACCCTGG
NTtmod1v4_10	CGGGCACTTGGCTGCCTTTG	ACAGCACACCTCAGCAACCCC
NTmyoD1_8	AAACGGCGGTCTGAAGAGCGAA	GCTTCACCCACCGGCACAGTT
NTnrap_3	TCAGCAGTGCAGTGAGAGGCTG	CTTGCTCCCCAGACGCCTTGT
NTtgfb3_5	CGTGGGTCGCTCCCCGAAAG	CCCATGCCTTGCTTCTCACAGGG
NTgadd45g_2	CGTCAGAGGGGAACAAGTCAGGT	CTGCATTGCTCTGTCTGCATGTGAA
NTcas3_8	CCCAGCAGCAGGCCCTTTAACC	GCAACACACTCAGGCCCGGAAAT
NTfoxo3_7	GAGCCCCAAAAGCGGCCAAG	GCCGCGCCAGAGCCATTG
NTcacnb1_10	GGTGGATCGGGCGTCTGGTG	GCCGGGCCACCTGTTCAG

List of primers used for real-time qPCR analyses of gene transcript analysis in the liver.		
Sequence (5'→3')		
Primer Name	Forward	Reverse
NTlpin2_4	AGGCTGGAGCTGACGCTCTGT	GGTGTCCGCCAGGAACCAAGATG
NTfabp10a_6	TTGAGGGCGGCAAACTGGTC	CGACCCAATGGTCATAGTCTCAACC
NTadipoR2_1	CAGCCAGACCACGGACAGACC	GGGAGCGGACAGACTTGCGG
NTpcdb1_5	CTGGTCCATCATCACATGACGTGC	ATACGGTGTAGACCATGGGTGTCG
NTscd_5	ATCGGCTAGTCTCCAGGGGTGG	GCCTGTCGCCTCTAGCTCGG
NTrxrgamma_9	CCGCCTCTTTTGCTCTTCACGG	CTTCGGGGCGCAAATGGTAGCG
NTmgs1_10	TGCACAGTGGCAACGAGAGGG	CGAGGCCTGTGAAGGGTCCG
NTfaxdc2_4	GTTTTCCGGGCTTGGGTCAGC	GCTTCCCCCAGCTTCCCTCG
NTabhd5_1	CGCCCTCTGGGCTCTCAACC	CACTCCTCCAGGGCTGCCAC
NTlss_3	CGGAGGGTTTGACTGATCTGACAGC	GGTCTGTGGCTGGCTCCGTC
NTelovl6_4	TGCTGTGGTCACTCACCTTGC	AGTGTACCCAGCTCTGGTGCTT

Appendix 4. A genetic linkage map of salt-tolerant tilapia constructed using all offspring.

There are 22 linkage groups named with LG1-22. The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers (see Appendix 1).

Appendix 5. A genetic linkage map of salt-tolerant tilapia constructed using only male offspring.

There are 22 linkage groups named with LG1-22. The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers (see Appendix 1).

There are 22 linkage groups named with LG1-22. The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers. (See Appendix 1).

Appendix 7. A genetic linkage map of salt-tolerant tilapia constructed using male and female offspring.

There are 22 linkage groups named with LG1-22. The number of the left is the distance in centiMorgans (cM). The labelling on the right are the names of the microsatellite DNA markers (see Appendix 1).

Appendix 8. QTL profile for growth traits and sex determination in all offspring.

Body weight – BW, Total length – TL, Sex determination – SD. The dotted line represents the threshold of genome-wide significance of $p < 0.05$.

Appendix 9. QTL profile for growth traits in male offspring.

Body weight – BW, Total length – TL, Body thickness – BT. The dotted line represents the threshold of genome-wide significance of $p < 0.05$

Appendix 10. QTL profile for growth traits in female offspring.

Body weight – BW, Total length – TL, Standard length – SL, Body thickness – BT. The dotted line represents the threshold of genome-wide significance of $p < 0.05$.

Appendix 11. Differentially expressed genes in the muscle in selected KEGG pathways.

Contig Name	NCBI Accession no	Description	Fold change of M_{fast} vs M_{slow}	Ko04390 Hippo signaling pathway	Ko04150 mTOR signaling pathway	Ko04151 PI3K-Akt signaling pathway	Ko04012 ErbB signaling pathway	Ko04010 MAPK signaling pathway	Muscle contraction	Ko04020 Calcium signaling pathway	Ko04152 AMPK signaling pathway	Ko03320 PPAR signaling pathway	Ko04144 Endocytosis	Ko01110 Biosynthesis of secondary metabolites	Fatty acid biosynthesis and	Unsaturated fatty acid biosynthesis	Alpha linoleic and linoleic acid metabolism ^a	Ko01212 Fatty acid metabolism	Ko00120 Primary bile acid	Ko04910 Insulin signaling pathway	Ko04922 Glucagon signaling pathway
c170799_g1_i5	XM 003440450.2	mothers against decapentaplegic homolog 3-like	-2.72																		
Contig10074	XM 003450805.2	calmodulin-like	3.57																		
Contig10093	XM 003450789.2	transaldolase-like	-2.59																		
		induced myeloid leukemia cell differentiation protein																			
Contig10348	XM 003450269.2	Mcl-1 homolog	-2.27																		
Contig10400	XM 003450147.2	forkhead box protein O1-A-like	-3.12																		
Contig10626	XM 003449867.2	sorting nexin-3-like	-2.38																		
Contig10739	XM 005454407.1	stromal interaction molecule 2-like	2.42																		
Contig11083	XM 003448710.2	S-adenosylmethionine synthase isoform type-1-like	2.13																		
Contig11131	XM 003448737.2	charged multivesicular body protein 6-like	-5.26																		
Contig11145	XM 003448369.2	cytochrome c1 heme protein mitochondrial-like	-2.03																		
Contig1120	XM 005460514.1	ribosomal protein S6 kinase alpha-6-like	-3.71																		
Contig11261	XM 003447840.2	M-phase inducer phosphatase 2-like	-2.10																		
		cyclic AMP-responsive element-binding protein 3-like																			
Contig11872	XM 003446148.2	protein 3-like	-3.07																		
Contig12096	XM 003444967.2	adiponectin receptor protein 2-like	-2.73																		
Contig1244	XM 003441471.2	GTP-binding protein Rheb-like	-2.17																		
Contig12561	XM 003443806.2	cytosolic 5'-nucleotidase 1A-like	2.46																		
		succinate dehydrogenase [ubiquinone] flavoprotein																			
Contig12610	XM 003443687.2	subunit mitochondrial-like	-3.04																		
Contig12809	XM 005450239.1	UPF0361 protein C3orf37 homolog	-4.80																		
Contig12949	XM 003440900.2	insulin receptor substrate 2-like	-3.37																		
Contig1309	XM 003449328.2	apolipoprotein A-I-like	4.12																		
Contig13287	XM 003439414.2	vacuolar protein sorting-associated protein 4A-like	-4.13																		
Contig13323	XM 003439386.2	dihydrolipoyl dehydrogenase mitochondrial-like	-2.15																		
Contig13405	XM 005449070.1	transferrin receptor protein 1-like	-2.64																		
Contig13576	XM 005448710.1	rabenosyn-5-like	-2.03																		
Contig14227	XM 005479136.1	epidermal growth factor receptor substrate 15-like	-2.55																		
Contig14346	XM 005478937.1	transcription factor AP-1-like	-2.29																		
Contig14527	XM 005478607.1	programmed cell death 6-interacting protein-like	-4.43																		
Contig14612	XM 003444186.2	14-3-3 protein beta/alpha-like	-2.34																		
Contig14672	XM 003444763.2	charged multivesicular body protein 4c-like	-2.32																		
Contig14738	XM 003444904.2	phosphogluconate dehydrogenase (pgd)	-3.84																		
Contig15345	XM 005477032.1	protein phosphatase 1A-like	-2.31																		
Contig15375	XM 005476973.1	ADP-ribosylation factor 6-like	-2.55																		
Contig15466	XM 003447482.2	kinesin family member 5B (kif5b)	-3.47																		
Contig1572	XM 005472239.1	proline and serine-rich protein 1-like	-2.13																		

[illegible]

Contig	Accession	Description	E-value
Contig2078	XM 005458256.1	rab GTPase-binding effector protein 1-like medium-chain specific acyl-CoA dehydrogenase	-3.71 -2.12
Contig2277	XM 003448445.2	mitochondrial-like	-2.12
Contig2466	XM 003443884.2	endoplasmic-like	-2.44
Contig2590	XM 003441528.2	mitogen-activated protein kinase 14 (mapk14)	-2.28
Contig2649	XM 003456076.2	ras GTPase-activating protein 1-like	-2.13
Contig2893	XM 005467076.1	E3 ubiquitin-protein ligase NEDD4-like	-2.05
Contig2971	XM 005475936.1	collagen alpha-2(VI) chain-like	2.02
Contig3176	XM 005454466.1	phosphatidylinositol 3-kinase regulatory subunit alpha-like	-3.80
Contig3248	XM 003446819.2	fructose-bisphosphate aldolase C-B-like	-2.10
Contig3387	XM 003448352.2	mannose-1-phosphate guanyltransferase beta-like	-6.50
Contig3417	XM 005471462.1	eukaryotic elongation factor-2 kinase (eef2k)	2.22
Contig3454	XM 003443999.2	bisphosphoglycerate mutase-like	2.01
Contig3512	XM 003449623.2	disabled homolog 2-like	-2.34
Contig3540	XM 003459358.2	chromatin modification-related protein eaf-1-like	2.45
Contig3647	XM 003443482.2	carboxymethylenebutenolidase homolog	-3.49
Contig3909	XM 003458877.2	very long-chain specific acyl-CoA dehydrogenase	-2.99
Contig3989	XM 005448306.1	mitochondrial-like	-2.01
Contig401	XM 003444360.2	mitogen-activated protein kinase 7-like	-4.89
Contig4019	XM 005459502.1	eukaryotic translation initiation factor 4E-binding protein 1-like	-2.34
Contig4062	XM 003438276.2	golgi-specific brefeldin A-resistance guanine nucleotide exchange factor 1-like	-9.09
Contig4087	XM 003440011.2	serine/threonine-protein phosphatase 2A 55 kDa	-2.00
Contig4143	XM 005459790.1	regulatory subunit B delta isoform-like	-2.73
Contig4284	XM 003450521.2	MAP kinase-interacting serine/threonine-protein kinase 1-like	-3.56
Contig4288	XM 003448208.2	tyrosine-protein kinase JAK2-like	-2.63
Contig4466	XM 003453869.2	transketolase (TK)	-3.33
Contig4630	XM 003450210.2	ras-related protein Rab-14-like	-2.45
Contig4657	XM 005468189.1	NEDD4-like E3 ubiquitin-protein ligase WWP1-like	4.69
Contig4774	XM 005453265.1	peroxisome proliferator-activated receptor gamma coactivator 1-alpha-like	-4.24
Contig4951	XM 003452341.2	thrombospondin-1b (thbs1b)	-2.09
Contig5233	XM 005473753.1	tyrosine-protein kinase JAK1-like	3.47
Contig5397	XM 005477940.1	protein phosphatase 1 regulatory subunit 3C-B-like	-2.01
Contig5537	XM 005474544.1	ADP-ribosylation factor guanine nucleotide-exchange factor 2 (brefeldin A-inhibited) (arfgef2)	-2.76
Contig5737	XM 005451611.1	acetyl-coenzyme A transporter 1-like	2.21
Contig5773	XM 003458284.2	voltage-dependent calcium channel subunit alpha-2/delta-1-like	-2.81
Contig592	XM 005448204.1	glucose-6-phosphate isomerase-like	2.64
		cytochrome c oxidase subunit 6B1-like	

[illegible]

[illegible]

Appendix 12. Differentially expressed genes in the liver of selected KEGG pathways.

Contig name	NCBI accession no.	Description	Fold change of L _{fast} vs L _{slow}	ko01040 Biosynthesis of unsaturated fatty acids	ko00061 Fatty acid biosynthesis	ko00062 Fatty acid elongation	ko03320 PPAR signaling pathway	ko04144 Endocytosis	ko01212 Fatty acid metabolism	ko00592 alpha-Linolenic acid metabolism	ko00591 Linoleic acid metabolism	ko01110 Biosynthesis of secondary metabolites	ko00100 Steroid biosynthesis	ko04910 Insulin signaling pathway	ko04922 Glucagon signaling pathway	ko00120 Primary bile acid biosynthesis	ko04390 Hippo signaling pathway	ko04150 mTOR signaling pathway	ko04151 PI3K-Akt signaling pathway	ko04012 ErbB signaling pathway	ko04010 MAPK signaling pathway	Muscle contraction	ko04020 Calcium signaling pathway	ko04152 AMPK signaling pathway
c171334_g1_i1	XP_005926249.1	glucose-6-phosphate 1-dehydrogenase-like isoform X1 [Haplochromis burtoni]	-2.33																					
Contig10212	XM_005455438.1	glyceraldehyde-3-phosphate dehydrogenase-like	2.65																					
Contig10262	XM_005455325.1	dehydrololichyl diphosphate synthase-like	5.25																					
Contig10278	XM_005455301.1	mitogen-activated protein kinase kinase kinase 5-like	-3.17																					
Contig10450	XM_003450035.2	dual specificity mitogen-activated protein kinase kinase 6-like	-10.85																					
Contig1047	XM_005477039.1	hypoxia-inducible factor 1-alpha-like	-4.18																					
Contig10529	XM_003449961.2	dual specificity protein phosphatase 5-like	-16.85																					
Contig1054	XM_003454966.2	adenylate kinase isoenzyme 1-like	2.11																					
Contig10549	XM_003449925.2	acyl-CoA synthetase short-chain family member 1 (acss1)	-3.05																					
Contig10639	XM_003449807.2	acetyl-CoA acetyltransferase mitochondrial-like	2.66																					
Contig10847	XM_005454185.1	hepatocyte growth factor receptor-like	3.02																					
Contig11141	XM_003448375.2	phosphoenolpyruvate carboxykinase cytosolic [GTP]-like	-6.11																					
Contig11143	XM_003448373.2	charged multivesicular body protein 4b-like	-5.12																					
Contig11481	XM_003447377.2	mitogen-activated protein kinase kinase kinase MLT-like	-5.75																					
Contig11655	XM_005452595.1	sorting nexin-12-like	-3.06																					
Contig12096	XM_003444967.2	adiponectin receptor protein 2-like	2.21																					
Contig12345	XM_005451182.1	serine--pyruvate aminotransferase mitochondrial-like	3.19																					
Contig12360	XM_005451140.1	guanine nucleotide-binding protein G(I)/G(S)/G(O) subunit gamma-12-like	2.77																					

Contig12365	XM_003444246.2	carnitine O-palmitoyltransferase 2 mitochondrial-like low density lipoprotein receptor adapter protein 1-B- like	-2.36
Contig1259	XM_003443813.2		2.97
Contig1276	XM_003441800.2	DNA damage-inducible transcript 4 protein-like succinyl-CoA ligase [GDP-forming] subunit beta	5.11
Contig12793	XM_003442757.2	mitochondrial-like	-3.22
Contig12802	XM_005450255.1	protein kinase C delta type-like	-2.52
Contig12838	XM_003442737.2	bcl-2-like protein 1-like	3.06
Contig1309	XM_003449328.2	apolipoprotein A-I-like	2.24
Contig13090	XM_003440779.2	connective tissue growth factor-like	-3.96
Contig13119	XM_003440824.2	vascular endothelial growth factor A-A-like	-2.39
Contig13166	XM_003439499.2	hydroxymethylglutaryl-CoA synthase cytoplasmic-like solute carrier family 27 (fatty acid transporter) member 6 (slc27a6)	-74.32
Contig13206	XM_003439467.2		4.44
Contig13340	XM_003439380.2	tyrosine aminotransferase (tat)	-2.18
Contig13423	XM_005449040.1	decaprenyl-diphosphate synthase subunit 1-like	-3.29
Contig13453	XM_005448983.1	transcriptional regulator Myc-like	-4.32
Contig1359	XM_005449800.1	cholesterol 24-hydroxylase-like segment polarity protein dishevelled homolog DVL-1- like	2.37
Contig13650	XM_005448547.1		-2.95
Contig13912	XM_003438531.2	aconitate hydratase mitochondrial-like	-3.22
Contig13939	XM_003438678.2	phosphatidylethanolamine N-methyltransferase-like isocitrate dehydrogenase [NAD] subunit alpha	3.08
Contig14045	XM_003437590.2	mitochondrial-like	-2.70
Contig14078	XM_005447711.1	glutamyl-prolyl-tRNA synthetase (eprs)	-2.64
Contig1413	XM_003452209.2	phosphoglucomutase 2 (pgm2)	-8.29
Contig14140	XM_003437654.2	pyruvate kinase muscle (pkm)	-3.72
Contig14256	XM_005479079.1	retinoic acid receptor RXR-gamma-B-like phosphatidylinositol 345-trisphosphate 3-phosphatase and dual-specificity protein phosphatase PTEN-like	3.43
Contig1443	XM_005474212.1		-3.02
Contig14457	XM_005478744.1	endophilin-A2-like	-2.57
Contig1447	XM_003442669.2	cytochrome b-c1 complex subunit 1 mitochondrial-like	-2.19
Contig14527	XM_005478607.1	programmed cell death 6-interacting protein-like	-2.34

[illegible]

[illegible]

Contig19412	XM_003441367.2	MAP kinase-activated protein kinase 2-like	-2.83
Contig19495	XM_003441414.2	tubulin polyglutamylase complex subunit 2-like	5.27
Contig19527	XM_003442550.2	nucleoside diphosphate kinase A-like	65.69
Contig19579	XM_003442589.2	dual specificity mitogen-activated protein kinase kinase 4-like	-2.83
Contig19605	XM_003442523.2	uncharacterized LOC100689736	3.42
Contig19682	XM_003442027.2	ATP-citrate synthase-like	-5.62
Contig1979	XM_003450130.2	suppressor of cytokine signaling 3-like	-8.48
		dihydrolipoilysine-residue succinyltransferase	
		component of 2-oxoglutarate dehydrogenase complex	
Contig1983	XM_003453071.2	mitochondrial-like	2.87
Contig19849	XM_003446410.2	aldehyde dehydrogenase family 16 member A1-like	2.63
		catenin (cadherin-associated protein) alpha 1 102kDa	
Contig20160	XM_003445854.2	(ctnnal)	-2.79
		cyclic AMP-responsive element-binding protein 3-like	
Contig2020	XM_003451315.2	protein 3-B-like	3.92
Contig2031	XM_003444726.2	EH-domain containing 1 (ehd1)	-3.88
Contig20497	XM_003443441.2	eukaryotic translation initiation factor 4E-1A-like	3.87
Contig20500	XM_003443399.2	elongation of very long chain fatty acids protein 6-like	10.85
Contig2073	XM_003448218.2	mitogen-activated protein kinase 14A-like	2.75
Contig2085	XM_003441456.2	ras-related protein Rab-7a-like	-3.33
Contig2129	XM_005449714.1	ADP-ribosylation factor 6-like	-2.22
Contig2193	XM_003457931.2	hydroxyacid oxidase 1-like	3.34
Contig2194	XM_005459781.1	long-chain-fatty-acid--CoA ligase ACSBG2-like	9.97
		medium-chain specific acyl-CoA dehydrogenase	
Contig2277	XM_003448445.2	mitochondrial-like	2.75
Contig2501	XM_003437821.2	signal transducing adapter molecule 1-like	-2.05
Contig2590	XM_003441528.2	mitogen-activated protein kinase 14 (mapk14)	-3.28
Contig2622	XM_005452654.1	vitronectin-like	3.38
Contig2649	XM_003456076.2	ras GTPase-activating protein 1-like	-3.40
Contig2686	XM_003453504.2	frizzled-6-like	2.09
Contig2709	XM_003451859.2	sterol O-acyltransferase 2-like	2.87
Contig2712	XM_005479143.1	sterol O-acyltransferase 1 (soat1)	-14.51

[illegible]

Contig4103	XM_003441491.2	25-hydroxyvitamin D-1 alpha hydroxylase mitochondrial-like	2.68
Contig4120	XM_003448158.2	glucose-6-phosphate dehydrogenase (g6pdh)	2.87
Contig4143	XM_005459790.1	tyrosine-protein kinase JAK2-like	2.37
Contig4231	XM_003446281.2	cytosolic 5'-nucleotidase 1A-like	-2.60
Contig4263	XM_005477623.1	low density lipoprotein receptor adapter protein 1-B-like	-3.25
Contig4281	XM_005472037.1	phospholipase D1-like	-2.61
Contig4458	XM_003438089.2	7-alpha-hydroxycholest-4-en-3-one 12-alpha-hydroxylase-like	30.16
Contig4611	XM_003448541.2	protein kinase C iota type-like	-3.75
Contig463	XM_003440857.2	acyl-CoA-binding protein-like	2.58
Contig4636	XM_003445449.2	ornithine decarboxylase-like	-2.24
Contig4668	XM_003453185.2	glycerol-3-phosphate dehydrogenase 1-like protein-like	3.28
Contig4724	XM_003441181.2	enoyl-CoA hydratase mitochondrial-like	3.22
Contig4729	XM_005475399.1	interleukin-1 receptor type 1-like	2.53
Contig4811	XM_003445698.2	IST1 homolog	-15.22
Contig4951	XM_003452341.2	tyrosine-protein kinase JAK1-like	-2.25
Contig4964	XM_003451333.2	lipoamide acyltransferase component of branched-chain alpha-keto acid dehydrogenase complex mitochondrial-like	-3.91
Contig4977	XM_005459928.1	arf-GAP with coiled-coil ANK repeat and PH domain-containing protein 2-like	-3.28
Contig5039	XM_003440688.2	calmodulin-like	-6.56
Contig5100	XM_003447604.2	purine nucleoside phosphorylase-like	2.20
Contig5193	XM_005464319.1	cyclic AMP-dependent transcription factor ATF-6 beta-like	2.73
Contig5215	XM_005451265.1	aldehyde dehydrogenase family 9 member A1-like	50.83
Contig5230	XM_003448636.2	ADP-ribosylation factor 1-like	-2.45
Contig5248	XM_005477098.1	DNA-binding protein inhibitor ID-2-like	-2.05
Contig5285	XM_003441339.2	cyclin-dependent kinase 4-like	-4.52
Contig5286	XM_003441785.2	short/branched chain specific acyl-CoA dehydrogenase mitochondrial-like	2.38
Contig5312	XM_003447521.2	catalase-like (catalase)	3.14

[illegible]

Contig8054	XM_003455828.2	phosphoglycerate mutase (pgm)	-3.33
Contig821	XM_003449972.2	very-long-chain enoyl-CoA reductase-like	-2.84
Contig824	XM_003456565.2	ras-related C3 botulinum toxin substrate 2-like	-3.49
Contig839	XM_003444047.2	fatty acid-binding protein heart-like	5.46
Contig8441	XM_005459447.1	epsin-2-like	2.31
Contig8486	XM_005459349.1	inositol-trisphosphate 3-kinase C-like	-8.49
Contig8500	XM_005459330.1	transcription factor RelB-like	-5.44
Contig8600	XM_003454562.2	aminoacylase-1A-like	6.75
Contig8853	XM_005458506.1	adapter molecule crk-like	-3.31
Contig8867	XM_003453729.2	lanosterol oxidase-like	-16.88
Contig8924	XM_005458349.1	phosphatidylserine decarboxylase proenzyme-like	3.04
Contig897	XM_003440410.2	isocitrate dehydrogenase [NADP] mitochondrial-like succinate dehydrogenase [ubiquinone] iron-sulfur subunit mitochondrial-like	-2.40
Contig907	XM_005454102.1		-2.17
Contig9144	XM_003452895.2	glycogen synthase kinase-3 beta-like	-2.20
Contig9150	XM_005457855.1	serine/threonine-protein kinase SIK2-like	-3.02
Contig9388	XM_003452465.2	lanosterol 14-alpha demethylase-like	-9.16
Contig957	XM_003446083.2	growth arrest and DNA damage-inducible protein GADD45 gamma-like	2.25
Contig9626	XM_005456749.1	PTK2 protein tyrosine kinase 2 (ptk2)	2.30
Contig9737	XM_003451930.2	eukaryotic translation initiation factor 4E-like	-2.73
Contig9861	XM_003451503.2	mothers against decapentaplegic homolog 7-like voltage-dependent anion-selective channel protein 1- like	-3.84
Contig9934	XM_003451434.2		-3.79
gil525343448	NM_001279671.1	heat shock protein 70	-2.82
gil525344164	NM_001279727.1	glucose transporter (glut1)	-2.24
gil525344577	NM_001279503.1	insulin-like growth factor 1 (igf-i) glyceraldehyde 3-phosphate dehydrogenase testis- specific-like	3.14
gil525345082	NM_001279552.1		-7.98
gil542171172	XM_003442250.2	diphosphomevalonate decarboxylase-like	-16.46
gil542171265	XM_003442244.2	cytochrome c oxidase subunit 4 isoform 1 mitochondrial-like	-2.03

[illegible]

[illegible]

[illegible]

Appendix 13. List of primers used for gene expression studies of candidate genes in muscle and liver of fast and slow- growing fishes.

	Sequence (5'→3')	
	Forward	Reverse
Sparc	ACC GAA GTG GGG GCC AAT CC	GTG GAG GGG TCC TGG CAC AC
FoxK1	CGG GGT CAT CGG GGT ATC GC	ACT CCG CTC GCT GCT CAC TC
Smad3	GAG GCA GTC TAC CAA CTC ACC AGG	CAC TAG AGC AGT GGA TGC TGG GG
Usp38	TCT TGG AAG CGT CTC GGC CC	GTG TCG GCT GGG GAG ACA GC
Crot	TCT GCC GGA GCA GAC CTT TCA G	CAG AGG CGA ACG GAC GAA CTG
Fdps	CCC GCC CTC CCA GCT TAC C	GGG TTG CCC TCT CCG AGT CA
Sqle	CGC CAT GGC CGA GCC AAA TG	ATG CGA GCC TTC CCG GAC AC
Cyp7b1	TGC AGT TCG TCC GTC AGG AGC	ATG GAG GAC GAG CAC AGG CG
Impa1	GGT GGT GAC AGA GAT CGG GGC	CAT GTC AAC CGC TGC TGT GCC
Gss	CAA TGG CCT TGG CAG CAC CAG	AGG GGC TGC CTT GTC TCA TCA AG