Domestication in Aquaculture Fishes- Elucidating the Genetic Consequences in Nile Tilapia (*Oreochromis niloticus*)

Maria George Nayfa

March 2020

For the degree of Doctor of Philosophy

Centre of Sustainable and Tropical Fisheries in Aquaculture,

and

Centre of Tropical Bioinformatics and Molecular Biology,

James Cook University, QLD, Australia

STATEMENT OF ACCESS

I, Maria George Nayfa, author of this work, understand that James Cook University will make this thesis available for use within the university library and via the Australian Digital Thesis network for use elsewhere. I understand that as an unpublished work, a thesis has significant protection under the Copyright Act and I do not wish to place any further restriction on access to this work.

.....

Maria G. Nayfa

March 2020

DECLARATION

I declare that this thesis is my own original work and has not been submitted in any form for another degree or diploma at any university or other institution of tertiary education.

Information derived from the published or unpublished work of others has been acknowledged in the text as a list of references is give.

.....

Maria G. Nayfa

March 2020

STATEMENT OF THE CONTRIBUTION OF OTHERS

I was involved in the conceptualization and experimental design of all work presented in this thesis. I was responsible for data management, integrity, analysis, and interpretation. This included: pedigree analysis and correction (Chapter 2); linkage mapping construction and annotation (Chapter 3); quantitative trait locus analysis and genome-wide association analysis (Chapter 3); detecting signatures of selection (Chapter 4), and population genomic analysis (Chapter 5). I was the sole author for all written work throughout the five chapters of this thesis and am the lead author in all peer-reviewed scientific manuscripts derived from this work whether published, submitted, or in preparation. Specific co-author contributions for this thesis are outlined by chapter in Table 1.

Table 1. Contributions of those who aided in the design and execution of this thesis.

Chapter	Individual	Nature and extent of intellectual input
1	Maria Nayfa	Literature Review, writing, and editing
	Kyall Zenger	Supervision and editing
	Dean Jerry	Supervision and editing
	David Jones	Supervision and editing
	John Benzie	Supervision and editing
2	Maria Nayfa	Execution and management of project, animal sampling design, data analysis, funding, writing, and editing
	Kyall Zenger	Supervision, advice on pedigree analysis and reassignment, and editing
	Dean Jerry	Supervision, funding, and editing
	David Jones	Supervision, advice on pedigree analysis and reassignment, and editing
	John Benzie	Supervision, funding, and editing
	Curtis Lind	Project conception and initial animal sampling design, initiation of genome sequencing, funding, supervision, and editing
	Khairul Rizal Abu Bakar	Plating and sending DNA samples to Diversity Arrays for extraction and genotyping
3	Maria Nayfa	Execution and management of project, data analysis, funding, writing, and editing
	Kyall Zenger	Supervision, funding, advice on linkage mapping and QTLs, and editing
	Dean Jerry	Supervision and editing
	David Jones	Advice on linkage mapping, QTL mapping, and GWAS approaches, supervision, and editing
	John Benzie	Supervision, funding, and editing
	Curtis Lind	Initial sampling design for linkage mapping and QTL mapping families
	Khairul Rizal Abu Bakar	Plating DNA samples to be to Diversity Arrays for extraction and genotyping
4	Maria Nayfa	Project conception, design, execution, and management, data analysis, writing, and editing
	Kyall Zenger	Project design, advice on population genomics, supervision, and editing
	Dean Jerry	Supervision, project design, and editing
	David Jones	Supervision, advice on population genetics and outlier analysis, and editing
	John Benzie	Supervision, project design, funding, and editing
	Khairul Rizal Abu Bakar	Plating DNA samples to be to Diversity Arrays for extraction and genotyping
	Muhammad Said	Collecting and sampling wild Nile tilapia in Egypt
5	Maria Nayfa	Writing and editing
	Kyall Zenger	Supervision and editing
	Dean Jerry	Supervision and editing
	David Jones	Supervision and editing
	John Benzie	Supervision and editing

AFFILIATIONS

Table 2. Student and supervisory panel affiliations.

	Role	Affiliation
Maria G. Nayfa	PhD Candidate	a, b
David B. Jones	Supervisor	a, b
Curtis E. Lind	Supervisor	c, e
John A. H. Benzie	Supervisor	c, f
Dean R. Jerry	Supervisor	a, b, d
Kyall R. Zenger	Supervisor	a, b
Khairul Rizal Abu Bakar	Lab Technician	c
Muhammad Said	Sample Collection	g

a. Centre for Sustainable Tropical Fisheries and Aquaculture, College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia

b. Centre for Tropical Bioinformatics and Molecular Biology, College of Science and Engineering, James Cook University, Townsville, QLD 4811, Australia

c. WorldFish, Jalan Batu Maung, Bayan Lepas, 11960 Penang, Malaysia

d. Tropical Futures Institute, James Cook University, Singapore

e. Commonwealth Scientific and Industrial Research Organisation, Castray Esplanade, Battery Point TAS 7004, Australia

f. School of Biological, Earth and Environmental Sciences, University College Cork, Cork, Ireland

g. Aquaculture Department, Faculty of Fish Resources, Suez University, Egypt

FUNDING

Research Funding

- WorldFish Centre ($45,500 AUD; 2016-2018)
  - European Commission and the International Fund for Agricultural Development (IFAD) Grant Number 2000001539
  - The United States Agency for International Development (USAID)
  - The CGIAR Research Program on Livestock and Fish Agri-Food Systems (LIVESTOCK AND FISH)
  - The CGIAR Research Program on Fish Agri-Food Systems (FISH) led by WorldFish.
    - The program is supported by contributors to the CGIAR Trust Fund.
- A Fisheries Society of the British Isles Small Research Grant (£5,000; 2017)
- Internal Research Grants from James Cook University
  - HDR Enhancement Scheme ($2,500 AUD; 2016-2017)
  - Joint Research Training Grant ($4,600 AUD; 2018)

PhD Scholarship Funding

- Sir Keith Murdoch Fellowship ($40,000 USD; 2015)
- JCU Postgraduate Research Scholarship ($58,200 AUD per annum; Jan 2016-Feb 2019)

Conference and Workshop Funding

- International Symposium of Genetics in Aquaculture XIII ($500 AUD; 2018)
- WorldFish Centre Aquaculture in Africa Meeting and Workshop ($4,000 AUD; 2019)

ACKNOWLEDGEMENTS

We're all just stories in the end, and the coming pages unfurl to tell the tales of my doctoral research. What isn't told, are the ups and downs that also ensued, those stories remain with me and with those whose paths crossed mine throughout the process.

First and foremost, I would like to express my sincere gratitude to my family (Ellada Nayfa, Fr. George Nayfa, Christina Nayfa, Aristotelis Nayfa, Martha Nayfa, Anna Gilman, and Makrina Nayfa). Despite always being ready to put me in my place, there was never a time that you doubted me nor would you let me doubt myself. You have always supported whatever decisions I've made and paths I've taken without question. Without your unwavering love and support (and occasional editing of drafts) I wouldn't be where I am today. You guys are my heart and soul- thank you.

To my 313 girls, you've been with me since the start of my academic journey. Meley Woldeghebriel, Veronica Ciocanel, and Beteal Ashinne- knowing that you girls are always in my corner kept me going, and our yearly adventures (including taking each of you out to see a reef) steeled me for the months on end in front of my computer. You have been my rock, and no matter the distance and no matter the problem you would always be there. I love you girls and I cannot wait to see what our next chapter brings.

To my goddaughters, Wylen Kyriakakis and Sadie Maria Williams, your joy of discovery and learning inspires and reminds me to take joy and wonder in what I do. I love you both so much! Lina Miraziz Jung, Georgiana Williams, Tanya Taylor, Amber Clifford, Angus Hogg and Kaleigh Russell thank you for always being there offering sound advice, friendship, and support. Andreas Gondikas, next time you give me career advice I might just listen. Thank you for the

encouragement and philosophical discussions. You helped keep me centered and sane. Tony Carrick and Daniel Driscoll thank you for always being there when I needed you.

To my Australia-based (at least when I initially met you) colleagues who quickly became my Australian family. I don't even know where to begin. We've been through so much together and have shared experiences that will bond us for a lifetime. Through the ups and the downs, thank you for always making the ride worthwhile: Diego Ortiz, Chao-Yang Kuo, Sandra Infante, Roger Heurlimann,, Alejandra Hernandez, Catarina Silva, Floriaan Devlo-Delva, Maximilian Hirschfeld, Eike Steinig, Kelsea Miller, Marta Espinheira, Ashton Gainsford and Aurelie Moya.

Diana Pazmiño Jaramillo and Natalia Andrade Rodríguez, thank you for always being there to catch me and occasional knock some sense into me. Nicolás Younes Cárdenas and Estefania Erazo-Mera, we didn't know each other very well when you guys took me in but living with you was one of the best things that could have happened to me. I'm happy to say that you are forever stuck with me and gato now. Jessica Grimm and Alex Nôiba Leurquin thank you for always being there for me (and Thekla) without question and for the endless flow of wine, conversation, and good times. Robert Strite, we've come full circle starting our MSc degrees together and now finishing our PhDs together- it's been a wonderful journey with you. Heather Loxton thank you for being there when I needed a confidant. Katie Sambrook and Mavi Gabela – you two have been the most amazing study buddies anyone could ask for. You kept me on track and motivated at the end when I couldn't find the motivation myself. Katie, you made all of those horrible long PhD weekends on City Campus fun. You and César Herrera Acosta have also been amazing hill buddies, our long talks up Castle Hill became the highlight of my week. Keep going you two- we're all almost at the top of the hill!

To my AquaGen family, thank you for being amazing. We've built friendships and connections that will not disappear. Rose Komugisha and Monal Lal, the two of you always bring a smile to my face. Your joy is contagious, and I am thankful to be collaborating with you both. Melissa Joyce, Shannon Kjeldsen, Alyssa Budd, and Madie Cooper you amazing women have been there during my some of my most vulnerable moments and saw me through to the other side. Thank you for being wonderful friends and a stellar support system. I'm looking forward to crossing that stage with all of you next year! Adrien Marc and Johannes Boyke, thank you for all of the chats and puppy dates! From bouncing ideas off one another to whinging about scripts that wouldn't behave to teaching me so much about aquaculture, a field which was foreign to me when I first started, you've been the absolute best office mate anyone could ask for Jarrod Guppy. A special thank you to Hugh Guppy, you have been a ray of sunshine bringing joy everywhere you go. Nothing beats building forts and playing hide-and-seek with you in the office!

I'd like to extend a special thank you to Justine Goddard and Dr. Susheel Suvarna. You gave me the tools that I needed to finish my PhD journey.

And finally, to my supervisory panel- thank you for helping me discover my strengths and for coming on this journey with me. Kyall Zenger, you are the one who first introduced me to the world of genetics and aquaculture. Dean Jerry, you have always been there with a level head and sound advice when I needed it. I've learned so much from you. John Benzie, it's been wonderful having you as supervisor and mentor- you've taught me a great deal. David Jones- I cannot begin to express my gratitude for your guidance and support throughout my PhD. I have no words but thank you.

ABSTRACT

A growing human population combined with a higher per capita consumption of fish has resulted in a greater demand for finfish. Fisheries production has plateaued over the last 30 years; thus, improving productivity in aquaculture to meet future market demands is vital. In particular, demand has increased for tilapia, the second most important group of commercially farmed finfish globally, due to their hardiness, low feed requirements and environmental adaptability.

Despite its success in terrestrial systems, uptake of genetic improvement techniques, like selective breeding, have only recently gained momentum in aquaculture. In 2002, the Abbassa Strain (AS) of Nile tilapia (*Oreochromis niloticus*) was established by the WorldFish Center in Egypt. The AS originated from a combination of both wild and hatchery population founders, with the objective to increase harvest weight using a combination of between and within family selection. To date, the AS has experienced 3.8-7.0% improvement in growth per generation, a modest increase compared to the 7.1-15.0% increase observed in similar Nile tilapia selective breeding programs. As little is known about the genetic state of the AS, this difference in genetic gain highlighted the need to examine the accuracy of AS management practices and whether sufficient genetic diversity has been maintained within the line for selection to act upon.

This thesis is the first comprehensive genetic study of a non-salmonid tropical finfish with six overarching objectives. These are to determine:

i) the accuracy of pedigree traceability and the management of the AS;

ii) the current and ongoing genetic status of the AS;

iii) the genetic architecture of commercially important traits;

iv) whether signatures of selection can be detected in the current stock;

v) the extent of wild population structuring in Nile tilapia in Egypt; and

vi) the potential effects AS escapees may have within its currently farmed regions and across Egypt where it is intended to be disseminated.

Traditionally, genealogical data has been utilized to monitor inbreeding rates, relatedness, and co-ancestry within selective breeding programs. Errors within genealogical records are common and have been shown to be as high as 15% in terrestrial selective breeding programs, with little information available on aquatic breeding programs. These genealogical errors can lead to inaccurate calculation of breeding values, a reduction in genetic gain, and inaccurate estimates of inbreeding.

To date, the AS has been managed solely based on genealogical data. To assess the accuracy of these genealogical records, firstly, stringently filtered genome-wide SNPs (1,040) were used to test and correct parentage assignments; secondly, 6,163 SNPs were used to determine the level of genetic diversity, the pedigree genetic structure and the number of families present within this line. Inbreeding coefficients and founder contributions were calculated from two founding events for 11 generations of the AS using molecularly corrected pedigree records. On average, AS pedigree error rates were found to be 45.5% per generation and are considered to be one of the most likely contributing factors leading to the relatively low genetic gain observed within the program. Inbreeding levels remained below 1% per generation; however, over 84.3% of available genetic material within the AS can be attributed to only 34 founders. This indicates that founder contribution has been eroded within the AS, and that optimal founder contribution should be taken into consideration in future management strategies to conserve genetic diversity while attaining genetic gain.

To better understand the genomic effects of selective breeding and the genetic architecture of weight and sex in *O. niloticus,* genomic resources for the AS were developed. This study produced the first line-specific linkage map for the commercially important AS. This linkage map is one of the first population-based genetic linkage maps using small families (16 families ranging from 5-17 offspring: 136 individuals) and phase unknown data, demonstrating the viability of this method for map construction. Due to the atypical construction of this linkage map, independent maps were created based on the sex average, female, and male lines. A total of 2,399 markers were successfully mapped to a sex average map, 2,197 to the female map, and 2,125 to the male map. All maps and map orders were validated by the reference genome assembly, Orenil 1.1 (GenBank Assembly Accession: GCA_00188235.2).

Phenotypic data was then utilized to undertake quantitative trait locus (QTL) analysis and genome-wide association studies (GWAS) to determine regions of the genome associated with sex and weight. QTL analyses were conducted using the two largest phased and phenotyped mapping families (8 and 10 offspring) available. Putative QTLs, or those QTLs observed in both families at a LOD > 10, associated with sex were found in LGs 12 and 23 in the sex average map and LG 23 in the male map. Suggestive QTLs, or QTLs identified in only one family at a LOD > 10, associated with sex were identified in LGs 8 and 14 in all three maps. Suggestive sex QTLs were also identified in LGs 3, 12, 19, and 23 in the female map as well as in LG 12 in the male map. GWAS identified LG 23 as being associated with sex in all three maps. Although karyotyping of *O. niloticus* identified a male heterogametic sex determining (XX|XY) system, to date, no study, including the present one, has clearly assigned a linkage group to the sex chromosomes as both genetics and environment can trigger the mechanisms underlying sex determination in Nile tilapia. Nile tilapia's sex determining system is further complicated by its

readiness to hybridize with other tilapia species; including, *O. aureus* which accounts for approximately 10% of the AS genome. Considering this, LGs 3, 14, 12, and 19 may be associated with sex determination either in *O. aureus,* or in the reproductive interaction between *O. niloticus* and *O. aureus*. However, to unravel these associations additional and more targeted analyses are required.

Weight was found to be a polygenic trait in the AS, with suggestive QTLs identified in LGs 2, 3, 8, 13, 14, and 18 in all three maps; LGs 6, 7, 11, 16, 17, 19, and 20 in the sex average map; LGs 4, 5, 6, 12, and 19 in the female map; and LGs 4, 5, 17, and 20 in the male map. However, these suggestive QTLs were not supported in GWAS analyses. Such results indicate that weight is indeed a complex trait governed by many genes of small effects; however, additional genotype by phenotype studies with a higher density of markers and more individuals are necessary to dissect growth in tilapia.

Domestication in conjunction with targeted selective breeding has a greater potential for detrimental genetic consequences, including the loss of genetic diversity and changes in allele frequencies compared to wild populations. For the first time, genetic data (9,287 SNPs) was used to identify population structure and signatures of selection amongst the AS and eight wild populations of Nile tilapia along the Nile River, Egypt. Two major genetic clusters (captive and wild populations) were observed. Wild populations showed evidence of isolation-by-distance between brackish Nile Delta and upstream riverine populations. Despite this, only a few outliers were detected in pairwise comparisons of wild populations. Approximately 6.9% of SNPs were identified as outliers (1.9% balancing outliers; 5.0% diversifying outliers) between captive and wild populations, but a lack of localized clustering suggests that no genes of major effect were detected. Subsequently, individuals belonging to the AS were easily distinguishable from

individuals originating from wild populations, with a putative first-generation escapee being detected in the wild. The AS was also found to have retained high levels of genetic diversity (H_{o_All} = 0.21 ±0.01; H_{e_All} = 0.23 ±0.01) when compared to wild populations (H_{o_All} = 0.18 ±0.01; H_{e_All} = 0.17±0.01) despite 11 years of selective breeding. Additionally, 565 private SNPs were identified within the AS line, which in addition to increasing AS heterozygosity, adds support to the finding that introgression with *O. aureus* has occurred. Wild populations all exhibited different subsets of polymorphic loci per sampling location, indicating that hybridization with *O. aureus* may have also occurred in the wild.

As a body of work, this thesis has found that both pedigree errors and the incorporation of genetic material from the smaller growing *O. aureus* into the AS have likely contributed to the modest genetic gain observed with the program. Despite this, genetic diversity indices and inbreeding levels within the program indicate that the AS is salvageable. To enhance future selective breeding efforts, three line-specific linkage maps were constructed using small family data. Novel and previously identified QTLs associated with both sex and weight were detected within the study, suggesting that both traits are polygenic. However, given small sample sizes and evidence of hybridization, further studies are required to validate these QTLs and determine their relevance to the AS before genomic selection is pursued. The AS were genetically distinct from their wild Nile tilapia counterparts, with putative AS escapees easily detectable. Wild population structure indicated some structuring due to isolation-by-distance; however, few outlier loci were detected amongst wild populations indicating that there are either no strong selective forces acting throughout their environmental range or that there is sufficient gene flow among populations to counteract selection. Additionally, signals of potential hybridization with *O. aureus* were detected in wild *O. niloticus* populations. Therefore, it could be speculated that

the disseminating the AS throughout Egypt should not have detrimental effects to natural populations or the performance of the AS itself.

PUBLICATIONS

I. Derived from Thesis Research

NAYFA, M. G., JONES, D. B., LIND, C. E., BENZIE, J. A., JERRY, D. R. & ZENGER, K. R. 2020. Pipette and paper: Combining molecular and genealogical methods to assess a Nile tilapia (*Oreochromis niloticus*) breeding program. *Aquaculture***,** 523, 735171.

NAYFA, M. G., JONES, D. B., BENZIE, J. A., JERRY, D. R. & ZENGER, K. R. (submitted manuscript, 2020). Comparing genomic signatures of selection between the Abbassa Selection Line and eight natural populations of Nile tilapia (*Oreochromis niloticus*) in Egypt. *Frontiers in Genetics.*

NAYFA, M. G., JONES, D. B., BENZIE, J. A., JERRY, D. R. & ZENGER, K. R. (in prep.). Novel associations to sex and growth within the Abbassa Selection Line of Nile tilapia (*Oreochromis niloticus*).

II. Derived from External Projects and Collaborations

NAYFA, M. G. & ZENGER, K. R. 2016. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (Pinctada maxima). *Marine genomics,* 28**,** 99-106.

WRIGHT, R. M., MERA, H., KENKEL, C. D., NAYFA, M., BAY, L. K. & MATZ, M. V. 2019. Positive genetic associations among fitness traits support evolvability of a reef-building coral under multiple stressors. *Global Change Biology,* 25**,** 3294-3304.

LAL, M.M., WAQAIRATU, S.S., ZENGER, K.R., NAYFA, M.G., PICKERING, T.D., SINGH, A., SOUTHGATE, P.C. (submitted manuscript, 2020).The GIFT that keeps on giving/ A genetic audit of the Fijian Genetically Improved Farm Tilapia (GIFT) broodstock nucleus 20 years after introduction. *Aquaculture.*

III. Government Reports

LAL, M.M, ZENGER, K.R., NAYFA, M.G., SOUTHGATE, P.C., WAQAIRATU, S. 2018. Consultancy Report CPS 17-674: Genetic Audit of the GIFT Nile tilapia breeding nucleus at the Naduruloulou Research Station, Fiji Islands.

CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF APPENDICES

ABBREVIATIONS

Abbassa Strain of Nile tilapia	AS
Average Inbreeding Coefficient	F
Base Pair	bp
Bayes Factors	BF
Centimorgan	cM
Change in Inbreeding Coefficient	ΔF
Cross Pollinators Population	CP
Double Haploid Population	DH
Effective Number of Founders	f^e
Effective Population Size	N_e
Estimated Breeding Value	EBV
Expected Heterozygosity	H_e
Expected Number of Homozygous Genotypes	$\hat{F}^{ii}$
First Generation of Offspring	F1
Genomic Estimated Breeding Value	GEBV
Genetically Improved Farmed Tilapia	GIFT
Genomic Breeding Value	GBV
Genomic Estimated Breeding Value	GEBV
Genome-wide Association Study	GWAS
Genomic Selection	GS
Genotype x Environment	GxE

Hardy-Weinberg Equilibrium	HWE
Identity-by-State	IBS
Inbreeding Coefficient	F_{is}
Interval Mapping	IM
K-Nearest Neighbor	kNN
Kruskal-Wallis Analysis	KW
Linkage Disequilibrium Method	LDN_e
Linkage Group	LG
Logarithm of the Odds	LOD
Marker Assisted Selection	MAS
Markov Chain	MC
Maximum Likelihood	ML
Mendelian Inheritance	MI
Minor Allele Frequency	MAF
Multi-locus Heterozygosity	MLH
Multiple QTL Model	MQM
Nearest Neighbor Fit	N.N. Fit
Observed Heterozygosity	H_o
Parent Generation	F0
Polymorphism Information Context	PIC
Proportion of the total genetic variance	F_{st}
Quantile-Quantile Plots	QQ-plots
Quantitative Trait Locus	QTL

Randomly Amplified Polymorphic DNA	RAPD
Reference Allele	REF
Single Nucleotide Polymorphism	SNP
The Change in the True Number of Clusters	ΔK
Total Number of Founders	f
True Number of Clusters	K

CHAPTER 1: GENERAL INTRODUCTION

1.1 The Current State of Global Fisheries and Aquaculture

The global human population is estimated to be 7.8 billion, with this number projected to reach 10 billion by 2060 (Dawson and Johnson, 2017). Approximately 42% of the global population (3.3 billion) obtain 20% or more of their animal protein intake from fish (FAO, 2019a). Over the next decade, both an increase in global population and an increase in *per capita* consumption due to increased wealth, particularly in Western countries, will result in an increase of approximately 1.2% per year in fish consumption (FAO, 2019a, Little et al., 2016). At present, 47% of food fish production is met by fisheries and 53% by aquaculture (FAO, 2019a), but as demand for fish protein continues to increase, these production ratios are expected to change towards higher aquaculture production.

Fisheries production has plateaued at approximately 90-95 million tonnes per annum over the past 30 years, with no foreseeable increase in production (FAO, 2019a). Thus, improving productivity in aquaculture is vital if future global consumption demands are to be met. Aquaculture production experienced a 6% increase per annum on average between 2001-2016, with production projected to continue to grow, but at a slower rate with some projections estimating only a 1.9% increase per annum (FAO, 2019a, Msangi et al., 2013). However, this does not have to be the case. Genetic improvement techniques can be used to further boost productivity in aquaculture.

Despite its ubiquity and success in terrestrial systems, genetic improvement techniques such as selective breeding are underutilized by aquaculture with most species obtained from either the wild or hatchery facilities in the early stages of domestication (Gjedrem and Baranski, 2010a,

Gjedrem et al., 2012, Gratacap et al., 2019). Given this, aquatic species have considerable genetic variation available which increases the potential for selective breeding programs to improve commercially important traits (Gratacap et al., 2019). A well-planned and managed selective breeding program can often obtain a 10-12.5% increase in productivity per generation for an aquatic species over the first few generations of selection (FAO, 2019a, Gjedrem et al., 2012). In fact, if selective breeding programs were established for all farmed aquatic species, the resulting increase in production could meet the projected increase in demand for fish with ease and require little extra feed, land, water or other inputs (FAO, 2019a, Gjedrem et al., 2012).

1.2 Selective Breeding

Selective breeding is directed evolution, where fitness is determined by a breeder rather than nature (Hill, 2001). Breeders use the natural genetic variation within a population to identify and mate individuals which exhibit traits of interest: for example, size (Gutierrez et al., 2015, Janssen et al., 2017), colour (Gjedrem and Rye, 2018, Zheng et al., 2013), or disease resistance (Houston, 2017, Moss et al., 2012). Aquaculture selective breeding programs more commonly than not use a closed nucleus mating system, in which no new genetic material is introduced into the line (Gjedrem and Akavaforsk, 2005).

In a closed breeding system, animals exhibiting favourable traits are mated to produce offspring with an increased frequency of desirable phenotypes. To accumulate genetic gain and continue improving the selection line, mate selection is repeated every generation with offspring of the previous generation. The long term success of a closed nucleus selective breeding program depends on a number of factors: including, trait heritability, selection intensity, additive genetic variance observed in founders, and the level of additive genetic variance maintained in each generation (Falconer et al., 1996, Loughnan et al., 2016). As the breeding program does not

introduce new germplasm into the line, individuals are becoming increasingly related with each subsequent generation and overall genetic diversity within the line is lost (Gjedrem and Baranski, 2010b, Pante et al., 2001).

Selective breeding balances rapid genetic gain against controlled decreases in inbreeding (Brisbane and Gibson, 1995). Within selective breeding programs, the rate of genetic gain can be adversely affected by a loss of genetic variation and increased inbreeding due to founder effects, genetic drift, high selection intensity, differential survival and parental contribution (Boudry et al., 2002, Frost et al., 2006, Lind et al., 2010). Aquatic programs are acutely vulnerable to these effects due to the life histories of these organisms: in particular, the high fecundity exhibited by aquatic species, which encourages the use of fewer founding individuals and results in less genetic diversity within the nucleus for selection to act upon (Gjedrem and Baranski, 2010c). To counteract this, it is essential to commence a breeding program with an adequate number of founders and to maintain a high effective population size (N_e) throughout the duration of the program (Gjedrem and Baranski, 2010a, Lind et al., 2012).

A substantial loss of genetic diversity within a selective breeding program can hamper productivity as it limits the amount of genetic variance available for selection (Falconer, 1960). Additionally, maintaining genetic diversity within a closed system is critical to accommodate current and future changes in production environments and if left unchecked, will result in increased homozygosity and deleterious fitness consequences associated with inbreeding (Pante et al., 2001). To counteract these effects, optimize the retention of genetic diversity and maximize genetic gains, selective breeding programs rely on the ability to trace pedigrees to assess relatedness and to manage family lines and ensure that consanguineous matings are avoided.

1.3 Nile Tilapia and Selective Breeding Programs

1.3.1 Nile Tilapia

Nile tilapia (*Oreochromis niloticus*) are the second most farmed fish globally after carp, accounting for 8.3% (4,525.4 thousand tonnes) of global finfish production (FAO, 2020). The Egyptian aquaculture industry for finfish is the largest in Africa (1,561.5 thousand tonnes; FAO, 2020), with *O. niloticus* accounting for 67.3% (1,051.5 thousand tonnes) of production (FAO, 2019b). Given its ubiquity in aquaculture production in addition to its robustness, short generation intervals, and tolerance for a wide range of environmental conditions (Rana, 1988, Avella et al., 1993, Shelton and Popma, 2006) *O. niloticus* are an excellent candidate for selective breeding.

1.3.1.2 Reproduction

Sexual maturity of *O. niloticus* can vary as it is a dynamic relationship among age, size, and environmental conditions (Shelton and Popma, 2006). *O. niloticus* has been recorded to reach sexual maturity at as small as 40g and as young as 5 months old (Rana, 1988, FAO, 2017). However, larger females tend to produce both larger and more eggs (Rana, 1988), with a 100g fish producing approximately 100 eggs per clutch and a 600-1,000g fish producing between 1,000-1,500 eggs per clutch (FAO, 2017). Larger eggs are associated with higher hatchability (higher hatching rates) and larger fry at hatching (Rana, 1988). Subsequently, these larger fry tend to exhibit a higher survival rate than their smaller counterparts (Rana, 1988).

O. niloticus are nesting substrate spawners whose females then incubate fertilized eggs, and later juveniles, in their mouths (Shelton and Popma, 2006). Juveniles progressively venture in

increasing distances and durations from the refuge of their mother's mouths until it is no longer necessary for their safety (Shelton and Popma, 2006).

1.3.1.2 Environmental Tolerance

Nile tilapia have the ability to withstand a wide range of environmental conditions: including, temperature, water quality, salinity, and acidity (Avella et al., 1993). Nile tilapia can tolerate temperatures between 12-42ºC, with temperatures beyond this range tolerated for short periods of time before mortality occurs (Balarin and Haller, 1982, Chervinski, 1982, Philippart and Ruwet, 1982). Although typically a freshwater fish, *O. niloticus* is able to survive in brackish water and tolerate salinities up to 30ppt (Avella et al., 1993, Kamal and Mair, 2005). Nile tilapia can acclimate to a pH of 4.0 (Dominguez et al., 2004) and can also withstand more alkaline conditions up to a pH 10 (Rebouças et al., 2016, Shelton and Popma, 2006). *O. niloticus* can also acclimate to fluctuations in ammonia toxicity (Shelton and Popma, 2006), and subsequently higher stocking densities which can result in higher ammonia concentrations through larger volumes of fish excretion (Salin and Williot, 1991).

1.3.2 Nile Tilapia Selective Breeding Programs

1.3.2.1 Genetically Improved Farmed Tilapia (GIFT) Strain

In 1988, the first Nile tilapia (*Oreochromis niloticus*) selective breeding program, the GIFT (genetically improved farmed tilapia) strain, was initiated in the Philippines and later transferred to Malaysia (Dey and Gupta, 2000). The GIFT strain was created from four wild Nile tilapia populations from Egypt, Ghana, Kenya, and Senegal, and four farmed populations from Israel, Singapore, Taiwan, and Thailand (Eknath et al., 1993, WorldFish, 2016). This breeding program

was highly successful and achieved 7.1-15.0 % improvement in growth per generation (Eknath and Acosta, 1998, Ponzoni et al., 2011).

To determine the dissemination potential of the GIFT strain, relative performances of growth, maturation, fecundity and hardiness were examined in different environments, i.e. genotype x environment (GxE) interactions (Eknath and Acosta, 1998). No significant GxE interactions were found, implying that in terms of performance, the GIFT strain would behave similarly in the environments tested, making it ideal for widespread distribution (Eknath and Acosta, 1998). The GIFT strain has been disseminated throughout 16 countries, mainly in Asia (Gupta and Acosta, 2004, Ponzoni et al., 2008, WorldFish, 2016). After dissemination, significant GxE interactions related to growth performance were observed in the GIFT strain when environmental conditions in different countries were examined (Agha et al., 2018). However, performance is not the only concern when disseminating a new strain.

There is a scarcity of studies examining the impacts of introducing a selected line of a native species developed in a non-native environment. However, research has shown that the introduction of captive individuals into a wild population can result in a phenomenon known as the Ryman-Laikre effect, where domesticated individuals can overwhelm wild populations and result in lower effective population sizes (Ansah et al., 2014, Ryman and Laikre, 1991). In addition, escapees, particularly from selectively bred lines, have also been shown to lower the fitness of wild populations (Yang et al., 2019) as demonstrated in Atlantic salmon, *Salmo salar* (Glover et al., 2013, McGinnity et al., 2003); European seabass, *Dicentrarchus labrax* (Toledo-Guedes et al., 2014); and Turbot, *Scophthalmus maximus* (Prado et al., 2018). While *O. niloticus* can now be found in areas all over the world, their traditional native range includes tropical and subtropical Africa without traversing past the Jordan Valley in the Middle East (Shelton and

Popma, 2006). Subsequently, protecting genetic diversity within these native ranges, being the origin and retainer of genetic diversity for the species, is of great importance. Another potential risk of introducing the GIFT from Asia back into its African home range is the risk of new disease-causing vectors and pathogens being introduced into unadapted African populations (Ansah et al., 2014). Thus, for fear of adverse impacts on native Nile tilapia germplasm, the GIFT strain was not considered viable for introduction back into Egypt (Ansah et al., 2014). However, the techniques utilized in the creation of the GIFT strain, labelled "GIFT Technology", were found to be transferable to other tilapia breeding programs (Gupta and Acosta, 2004): including, the Abbassa Strain of Nile tilapia in Egypt.

1.3.2.2 The Abbassa Strain of Nile Tilapia

In 2002, the Abbassa Strain of Nile tilapia (AS) was established by the WorldFish Center (formerly known as ICLARM) in Egypt (Ibrahim et al., 2013, Rezk et al., 2009). The AS originated from a combination of both wild populations (Aswan, Zawia, and Abbassa) and a hatchery population as founders in 2002. The objective of this line was to increase growth, measured as final harvest weight, using a combination of between and within family selection (Ibrahim et al., 2013, Rezk et al., 2009). Compared to the GIFT strain, the AS experienced a modest 3.8-7.0% improvement in growth per generation (Rezk et al., 2009). Despite this modest genetic gain, the AS still outperforms other commercially, non-selectively bred strains of Nile tilapia in Egypt, such as the Kafr El Shekh strain (Ibrahim et al., 2013). Here the AS outweighed this popular strain by 28% at harvest, regardless of stocking densities (Ibrahim et al., 2013). Additionally, while the females of both strains were smaller than their male counterparts, AS females grew at a comparable rate to males from the Kafr El Shekh strain (Ibrahim et al., 2013).

Increased growth rate within the first year is of particular interest in rural areas as tilapia grow quickly during warm summer months, but growth dramatically slows or halts in winter (Azaza et al., 2008, Platt and Hauser, 1978, Rezk et al., 2009). This type of growth pattern is not economical for small, rural fish farms who rely on earthen ponds, and must have their stocks reach market size before unfavourable winter conditions occur (Rezk et al., 2009). As such, breeding a strain of tilapia that exhibits a faster growth rate, resulting in larger fish at harvest and more animal product for the community, is highly desirable in Egypt.

1.4 Advancing Genetic Breeding Programs

Management of the AS based on phenotypic information alone has resulted in a 3.8-7.0% improvement in growth per generation (Rezk et al., 2009). However, a comparison of this modest genetic gain to similar Nile tilapia strains, like the GIFT strain which includes founders from Egyptian populations and almost double the genetic gain (7.1-15.0% genetic improvement per generation; Eknath and Acosta, 1998, Ponzoni et al., 2011), identifies an unexpected variation between the two strains. To improve production within the AS in future generations, the source of this variation needs to be understood; whether it arises from the AS founding population having limited diversity compared to the GIFT strain, or whether the program has not been optimally managed.

Once the source of this modest improvement is understood, either marker assisted selection (MAS) or genomic selection (GS) can be utilized to further improve animal selection within the AS. While both MAS and GS rely on the concept that genes associated with traits of interest will be in linkage disequilibrium with a minimum of one marker, otherwise known as “hitchhiking” (Goddard and Hayes, 2007, Smith and Haigh, 1974, Zenger et al., 2017), they differ in their execution. MAS works best with genes of major effect to aid animal selection (Arruda et al.,

2016); however, the effectiveness of this technique is limited when the trait is complex and controlled by many genes of smaller effect (Zenger et al., 2019). GS, in contrast, uses all available genome wide markers to provide a more accurate prediction of genetic merit for a trait, calculate genomic breeding values and streamline the selection process for complex traits of interest (Zenger et al., 2017).

1.5 Genomic Resources for Nile tilapia

1.5.1 Genome Assemblies and Linkage Maps

Although it is possible to conduct MAS and GS using unordered genome-wide markers, it is beneficial to utilize a genome assembly or a robust, high-density genetic linkage map to order markers. Marker order can be used to better understand the effects of selective breeding on the *O. niloticus* genome and the genetic architecture of specific traits, like sex or weight, through QTL mapping and GWAS (Du et al., 2016, Tsai et al., 2015). If a gene of major effect is detected, it can then be used to direct MAS; alternatively, if the trait is polygenic, a GS statistical model may be necessary to improve estimated breeding value (EBV) prediction accuracy (Zenger et al., 2019).

Currently, there are two genome assemblies (*O. niloticus* Orenil 1.1 and O_niloticus_UMD1) and five linkage maps published for *O. niloticus* (Conte et al., 2017, Guyon et al., 2012, Joshi et al., 2018, Kocher et al., 1998, Lee et al., 2005, NCBI, 2017, Palaiokostas et al., 2013). Both genome assemblies have been assembled to the chromosomal level; however, both have a substantial number of unplaced scaffolds (2,460-5,655; NCBI, 2017). Unplaced scaffolds can be problematic for molecular studies as genes of interest may be fractured or incorrectly annotated

(Baker, 2012, Denton et al., 2014). Linkage maps can be used to check marker placement and position of previously unmapped markers (Fierst, 2015).

There are currently five linkage maps available for *O. niloticus*, with most maps constructed from University of Sterling Stock and GIFT (or GIFT derived) stock (Table 1.1; Guyon et al., 2012, Joshi et al., 2018, Lee et al., 2005, Palaiokostas et al., 2013). The largest and most recent map constructed in 2018 from the Genomar Supreme Tilapia Strain, derived from the GIFT strain, consists of 40,186 SNPs mapped to 22 linkage groups (LG) which spans a total map length of 1,469.69 cM and has an average gap interval of 0.04 cM (Table 1.1; Joshi et al., 2018). Only the two most recent, and largest, linkage maps were constructed using single nucleotide polymorphism (SNP) data (Joshi et al., 2018, Palaiokostas et al., 2013). Unlike genome assemblies which determine physical distance of markers, linkage maps rely on rates of meiotic recombination between parents and offspring to determine the relative position of markers. These rates can vary greatly between species, populations, individuals and even genomic regions (Dukić et al., 2016). As such, it is imperative to create line-specific high-density genetic linkage maps for aquatic species.

Table 3.1 The available linkage maps for Nile tilapia (*Oreochromis niloticus*) with the following map metrics reported: map length (cM), number of markers, average interval, and marker type.

Tilapia Species	*O. niloticus* Strain	Map Length (cM)	Number of Markers	Marker Type	Average Interval	Authors & Year
O. niloticus	*Unknown*	701	62; 112	Microsatellites ; AFLP	--	(Kocher et al., 1998)
O. niloticus x *O. aureus*	University of Sterling Stock	1,311	525 21	Microsatellites Gene-based markers	2.4	(Lee et al., 2005)
O. niloticus	University of Sterling Stock & GIFT	34,084 cR 3500 937,310kb	1,358	Markers and radiation hybrid (RH) Map	--	(Guyon et al., 2012)
O. niloticus	University of Sterling Stock	1,176	3,802	SNPs	0.7	(Palaiokostas et al., 2013)
O. niloticus	Genomar Supreme Tilapia (derived from GIFT)	1,469.69	40,186	SNPs	0.04	(Joshi et al., 2018)

1.5.2 Quantitative Trait Loci (QTLs) and Genome-Wide Association Studies (GWAS)

Linkage maps for Nile tilapia have been used to detect both growth and sex related markers via QTL analyses and association analyses through genome-wide association studies (GWAS). While both methods rely on phenotypic data, QTL analyses have greater detection power when families with large numbers of offspring are available, whereas GWAS require large, heterogeneous, populations to resolve linkage disequilibrium in markers (Hayes, 2013, Korte and Farlow, 2013). Due to these variations in detection and strength, it can be beneficial to utilize both methods in conjunction with one another. Previous studies have used these the combination of these two methods to detect markers associated with weight in nine linkage groups in Nile tilapia and other tilapia species (Liu et al., 2014, Lin et al., 2016). Sex-linked markers have been associated with four linkage groups (Conte et al., 2017, Palaiokostas et al., 2015, Lee et al., 2003, Cáceres et al., 2019, Eshel et al., 2011). Despite these studies, both sex and weight in Nile tilapia appear to more complicated than first believed and there is still a great deal that we do not understand about the architecture of these complex traits (Baroiller et al., 2009, Cáceres et al., 2019, Conte et al., 2017, Eshel et al., 2011, Lee et al., 2003, Liu et al., 2014, Mank, 2008, Palaiokostas et al., 2015, Wang et al., 2019).

1.6 Comparing Domestic and Wild Populations of Nile Tilapia Throughout Egypt

Animals in selective breeding programs not only undergo selection for desirable traits like size (Argue et al., 2002, Eknath and Acosta, 1998), disease resistance (Moss et al., 2012, Robinson and Hayes, 2008) and colour (Hossain et al., 2011, Wan et al., 2017), but also adapt to a captive environment by displaying reduced antipredator behaviors and aggression (Johnsson et al., 1996, Robinson and Hayes, 2008). This targeted selection experienced by captive populations yields different genetic consequences to the natural selection their wild counterparts have undergone.

To understand these genetic consequences and identify signatures of selection, metrics relating to the genetic health of captive and wild populations of the same species can be compared (López et al., 2019, Simmons et al., 2006).

A clear distinction between wild and domestic populations has been observed in Atlantic Salmon, *Salmo salar* (Gutierrez et al., 2016) and gilthead sea bream, *Sparus aurata* (Cossu et al., 2019) due to founder effects, genetic drift, and the subsequent selection for domestic conditions and traits of interest. Domestic populations of Atlantic Salmon, *Salmo salar* (Bentsen and Thodesen, 2005), Pacific oyster, *Crassostrea gigas* (Zhong et al., 2017), and gilthead sea bream, and *Sparus aurata* (Cossu et al., 2019) have demonstrated lower genetic diversity and smaller effective population sizes. These strong differences in population structure make the detection of escapees and monitoring their consequences on wild populations possible. Escapees can have various effects on wild populations. For instance, escapees from domesticated lines in Atlantic salmon, *Salmo salar* (Glover et al., 2013, McGinnity et al., 2003); European seabass, *Dicentrarchus labrax* (Toledo-Guedes et al., 2014); and Turbot, *Scophthalmus maximus* (Prado et al., 2018) have been shown to lower the fitness of wild populations (Yang et al., 2019); whereas, other domesticated lines, like domestic rainbow trout (*Oncorhynchus mykiss*), have been shown to exhibit lower survival rates in the wild, with little to no effect on wild population genetics, due to their increased size and bolder foraging habits exposing them to higher predation (Biro et al., 2004).

In addition to identifying the potential genetic consequences of escapees, comparing wild and domestic populations can be used to identify signatures of selection. Hundreds of outlier markers have been detected between domestic and wild aquatic populations; including, brown trout *Salmo trutta* L. (Linløkken et al., 2017) and Atlantic salmon, *Salmo salar* L. (López et al., 2019).

These markers can then be associated with specific regions of the genome under selection and, if detailed genomic annotations are known, could be associated with specific genes (López et al., 2019, Marrano et al., 2018). In turn, these findings can not only be used to improve MAS and GS within the domestic line, but to screen founders for future programs for founders.

1.7 Thesis Overview

The objective of this thesis is to understand the genetic health, effects of domestication and demonstrate the importance of selective breeding program management practices for the selectively bred Abbassa Strain of Nile tilapia (*Oreochromis niloticus*), with reference to its intended dissemination throughout Egypt. In addition, this thesis also evaluates the possibility of incorporating advanced genetic breeding techniques into the already established breeding program to increase observed genetic gains by investigating the genetic architecture of two commercially important traits, sex and weight. This will be accomplished by i) comparing pedigree and molecular data generated from generations 9, 10 and 11 of the AS, ii) constructing a high-density genetic linkage map for the AS, iii) identifying QTLs and GWAS associated with sex and weight iv) exploring the genetic architecture of domestication, and v) identifying genomic variations among the AS and eight wild populations of *O. niloticus* in Egypt. This work has been presented in three subsequent data chapters outlining core investigations.

Chapter 2 utilizes both historic genealogical data and genome-wide markers to compare estimates in relatedness and effective population size within the breeding nucleus in order to correct genealogical records and identify the genetic consequences of pedigree errors on the AS.

Chapter 3 utilizes previously constructed genome assemblies of *O. niloticus* along with phase unknown, two generational, family lines of the AS to create a strain specific genetic linkage map

for the AS. This linkage map is then used to investigate the genetic architecture of sex and weight traits in the AS by identifying quantitative trait loci (QTLs) and conducting genome-wide marker association studies (GWAS).

Chapter 4 investigates the natural population structure of eight wild populations of Nile tilapia along a geographical gradient on the Nile River, Egypt, detect any differences in genetic diversity between these natural populations, and identify any signatures of selection among the AS and the natural populations.

CHAPTER 2: PIPETTE AND PAPER: COMBINING MOLECULAR AND GENEALOGICAL METHODS TO ASSESS A NILE TILAPIA (*Oreochromis niloticus*) BREEDING PROGRAM

2.1 Introduction

Aquaculture selective breeding programs employ a closed nucleus mating strategy whereby animals displaying sought-after characteristics are mated to produce next generation offspring with increased prevalence of desirable phenotypes. Offspring exhibiting high genetic merit for favorable traits are then usually chosen as candidate parents for the subsequent breeding cycle. The selective breeding process is replicated each succeeding generation in order to accumulate genetic gain within the breeding population. The long term success of these closed breeding systems is dependent on a number of factors: including, the heritability of a trait, the intensity of selection, the additive genetic variance observed in the founding population, and the amount of additive genetic variance maintained over subsequent generations (Falconer et al., 1996, Loughnan et al., 2016). If breeding practices are not properly managed, the number of animals with high relatedness will increase (Gjedrem and Baranski, 2010b), leading to a substantial loss of genetic diversity over subsequent generations (Pante et al., 2001). The maintenance of genetic diversity is critical to accommodate current and future changes in production environments, and if left unchecked, it can lead to inbreeding through increased homozygosity and deleterious fitness consequences associated with inbreeding depression (Pante et al., 2001). This loss of genetic diversity can also hamper progress within the selective breeding program as it limits the amount of genetic variance available for selection (Falconer, 1960).

Aquaculture selective breeding programs may be particularly vulnerable to both a loss of genetic variation and inbreeding due to founder effects, genetic drift, high selection intensity, differential survival and parental contribution (Boudry et al., 2002, Frost et al., 2006, Lind et al., 2010). This increased susceptibility to a loss of genetic variation and inbreeding is often due to the life histories of many aquatic animals: including, asynchronous spawning (Nguyen, 2016); larval sizes below minimum sizes for physically tagging individuals (Ouedraogo et al., 2014); and high fecundity, which encourage the use of fewer founding individuals within a selective breeding program (Gjedrem and Baranski, 2010c). To counteract these effects and maintain genetic diversity while maximizing genetic gains, aquaculture selective breeding programs rely on the ability to trace pedigrees to assess relatedness to manage family lines based on founders. As the selection program progresses, family lines share a greater and greater co-ancestry with one another and the potential for inbreeding increases (Gjedrem and Baranski, 2010b). Additionally, shared co-ancestry amongst families occurs more rapidly in selective breeding programs because the selection of individuals from each family is not random, and they likely contain favorable quantitative trait loci from the same few high performing founders (Sonesson et al., 2012).

Traditionally, genealogical data has been utilized to monitor inbreeding rates, relatedness, and co-ancestry within selective breeding programs; however, error rates in genealogical records have been shown to range from 1-15% within a terrestrial selective breeding program (Bovenhuis and Van Arendonk, 1991, Crawford et al., 1993, Sanders et al., 2006). Genealogical records in aquatic selective breeding programs have the potential to be more erroneous due to large family sizes and difficulties in retaining pedigree throughout the production cycle, particularly in juvenile stages. Genealogical errors lead to inaccurate estimated breeding values (EBVs), a reduction in genetic gain, and inaccurate estimates of inbreeding within the selective

breeding program (Banos et al., 2001, Israel and Weller, 2000). The degree to which these inaccuracies affect genetic gains is correlated with the percentage of errors and the length of the selective breeding program. In general, as genealogical errors within the population increase, the number of inaccuracies detected are expected to rise (Banos et al., 2001). In order to reduce these errors, accurate molecular data can be used to correct genealogical records, decrease inbreeding occurrences, and improve estimates of EBVs and genetic gains (Israel and Weller, 2000, Munoz et al., 2014, Visscher et al., 2002).

Considering the potential impact seed supply has on production systems, correctly managing nucleus breeding programs is critical to optimize genetic gain in production systems. With over 4,199,567 metric tons of farmed fish produced in 2014, tilapia are the mostly widely farmed fish globally (FAO, 2017). Tilapia is a fecund freshwater species with a short generation period and sexual maturity reached as young as six months of age (Duponchelle and Panfili, 1998). This robust fish is of particular importance in developing countries where it is grown to not only quickly and efficiently meet local protein requirements, but also help improve local job markets (FAO, 2017).

The Abbassa Strain (AS) was initially established by the WorldFish Center (formerly known as ICLARM) in Egypt and relied solely on genealogical data to calculate genetic diversity estimates. The AS originated from a combination of both wild (Aswan, Zawia, and Abbassa) and a hatchery population as founders in 2002, with the objective to increase growth rate using a combination of between and within family selection (Ibrahim et al., 2013, Rezk et al., 2009). This closed selective breeding line produces and maintains approximately 110 full-sibling families per one-year generation interval, with matings occurring once per year and each broodstock only contributing to a single mating season. At mating, one adult male and two adult

females are held together in a mating hapa. All mating hapas are labeled with a numbered to identify each family and placed into a single pond. Once one female spawns (i.e. is found to have eggs in her mouth), the male and the other female are removed from the mating hapa, leaving the spawned female to incubate the eggs. Offspring from a single family are then reared in three replicate hapas, each hapa containing 25–30 fry, and the remaining fry are kept in the mating hapa. All hapas are kept closely spaced in one large pond. Once fingerlings are large enough to be tagged with PIT tags, at approximately 8.5 g, the fish are then communally reared until harvest. At harvest, each fish is identified through PIT tags, sexed and weighed. Selection is based solely on the EBV for harvest weight.

The 3.8-7.0% improvement in growth per generation of the AS (Rezk et al., 2009) is much less than the 7.1-15% improvement per generation reported for the GIFT Strain (Eknath and Acosta, 1998, Ponzoni et al., 2011). This brought to question what factors that might explain this relatively slow rate of improvement (for instance, founder quality, pedigree quality, and overall management). For example, in the 4th generation of the AS, new germplasm (2,178 animals) from a sister, less intense, selective breeding line of Nile tilapia were introduced into the AS line. However, only 94 of these individuals were used as broodstock for the next generation. It is unknown how well these introduced individuals integrated into the AS over subsequent generations.

This study used genome-wide molecular information to correct pedigree records and determine the effects on genetic estimates, examine genetic diversity, infer genetic relationships, estimate founder contribution (including, one secondary founder introduction) and inbreeding rates, determine the factors that contributed to the relatively slow rate of improvement observed within

the AS, and finally, understand the current state of the strain to assist in determining future strategies for improvement.

2.2 Methods

2.2.1 Genealogical Analysis

Pedigree records of 28,781 individuals from generation 0 (founders) to generation 11 of the AS were provided by WorldFish. Each broodstock was only used in one breeding period and there was generational turnover each year. Genealogical analyses were conducted within PEDIG (Boichard, 2002) and ENDOG v4.8 (Gutiérrez and Goyache, 2005) using both the original pedigree and pedigree records corrected by molecular data for generations 10 and 11 (as described in section 2.2.3). Inbreeding was calculated utilizing the PEDIG program *meuw* (Boichard, 2002). The total number of founders (f) and the effective number of founders (f^e) were calculated via the PEDIG program *prob_orig* (Boichard, 2002). The percentage of each founder genome from both founding events (generation 0 and generation 4) that can be observed in subsequent generations were calculated using ENDOG v4.8 (Gutiérrez and Goyache, 2005). Analyses were conducted using all recorded individuals as well as only those records of individuals who contributed to the subsequent generation. Differences in results was negligible; therefore, only results from the dataset including all individuals are reported.

2.2.2 Molecular Analysis

2.2.2.1 Sampling, DNA extraction, and genotyping

To the date of these analyses, the AS has been running for 11 generations. Tissue samples were obtained from subset generations 9-11 as they were the only generations with tissue samples available [122 individuals from generation 9 (G9); 216 individuals from generation 10 (G10);

and 54 individuals from generation 11 (G11)]. DNA extractions, genotyping, and co-analysis were conducted by Diversity Arrays Technology (DArT) as described in Lind et al. (2017). To ensure genotype reproducibility, 20% random technical replicates were included internally at DArT.

DNA extractions and genotyping were conducted by Diversity Arrays Technology as described in Lind et al. (2017). In short, DArTseq™ employs a combination of complexity reduction methods and next generation sequencing in which two enzymes, a rare cutting restriction enzyme as well as a more frequently cutting enzyme (similar to double digest RAD sequencing, or ddRAD; Courtois et al., 2013, Kilian et al., 2012, Peterson et al., 2012, Raman et al., 2014, Von Mark et al., 2013). Library construction was optimized by first testing four restriction enzyme combinations (Pst-Hpall, Pstl-Sphl, Pstl-Msel, and Pstl-Mspl), where Pstl-Hpall proved to be the optimal enzyme combination for *O. niloticus* based on the number of markers detected as well as technical parameters (call rate, polymorphic inforamtion content, average read depth, average count ratios, and average reproducibility, Lind et al., 2017). Once an enzyme combination was selected, efficient sequencing selection and the implementation of Dartsoft14 in the KDCompute framework were utilized to finalize the library (Lind et al., 2017). To ensure genotype reproducibility, 20% random technical replicates were included internally at DArT.

2.2.2.2 SNP filtering

Prior to filtering, three individuals with greater than 30% missing data were removed, resulting in a total of 388 individuals (121 G9, 216 G10, and 51 G11) available for subsequent analyses. Single nucleotide polymorphism (SNP) data was then stringently filtered to ensure that only the highest quality and most informative markers were utilized for molecular analyses. A custom Python script, DartQC (https://github.com/esteinig/dartqc), was used to select single unique

SNPs within clone sequence tags from the dataset, silence genotype calls within each SNP that had less than five read counts, remove SNPs with a minor allele frequency (MAF) less than 0.01, remove SNPs with a call rate less than 0.8, and remove SNPs with an average replication statistic of less than 0.9 (Kjeldsen et al., 2019, Nayfa and Zenger, 2016).

Raw clone sequences from which SNPs were identified during the DArTseq process were annotated to the available genome assembly for *Oreochromis niloticus* (GenBank Assembly Accession: GCA_00188235.2; Orenil1.1) using a custom Perl script based on NCBI CGI BLAST interface with a 70% minimum sequence identity (Heller-Uszynska et al., 2011). As two linkage groups, LG 1 (Palaiokostas et al., 2013) and LG 23 (Eshel et al., 2012), are known to have sex-linked loci, SNP markers within either of these two linkage groups, or that were unassigned to a linkage group, were examined for Mendelian Inheritance ratios for sex linked markers, however, no sex linked markers were identified. A total of 6,163 high quality and informative SNPs were retained.

2.2.3 Pedigree Analysis

Pedigree records from generation 0 (founders) to generation 11 of the AS were collated from farm records. To test the accuracy of first order relatives in generations 9, 10, and 11 of the AS, comparisons were made between farm pedigree records and the parental assignments based on genotypic data. In order to identify and reduce the potential for miss-assignments in molecular pedigree analysis, a hierarchical pipeline was employed which utilized a random subset of 1,040 genome-wide SNPs. Firstly, this hierarchical approach utilized CERVUS v. 3.0, which employs a pairwise likelihood-based parental assignment method to test and identify true parents of individuals from each generation separately (Kalinowski et al., 2007). The AS has a yearly mating scheme consisting of a one male to two female ratio within a hapa; however, although

unlikely, the close proximity of hapas could results in mating between fish from different hapas via fish potentially jumping or sperm transferred between adjacent hapas. Therefore, respective paternity relationships were tested first using a paternity parentage analysis. Corrected paternal relationships were then utilized in a maternity parentage analysis. A polygamous paternal and monogamous maternal model was utilized with 5% genotyping error rates (Kalinowski et al., 2007). All parental assignments were then confirmed using a second maximum likelihood-based parental assignment program, COLONY v. 2.0.6.3, with a 5% genotyping error rate (Jones and Wang, 2010). This second program was not only used for confirmation, but for its added feature which allows the identification of family groups (i.e. parents and siblings). Only parental matches with greater than 95% probability of correct pair assignments in both Cervus and Colony were retained. Once reassignments were completed, further verification was undertaken using pairwise genetic relationships calculated in GCTA v.1.26.0 (Yang et al., 2011). GCTA v.1.26.0 establishes all individuals analyzed as the base population and defines relatedness so that the average relatedness between "unrelated" individuals is zero (Powell et al., 2010, Yang et al., 2011). As the presence of missing data, low marker MAF and inbreeding distorts the calculation of relatedness, the margin of variation from traditionally accepted relationships (for example, a 50% relatedness between full-siblings, 25% relatedness between half-siblings and a 50% relatedness between an offspring and its parents) was scaled per individual based on the relatedness calculated to self. Finally, for the family groupings where both parents were genotyped, Mendelian Inheritance errors were also confirmed to be less than 3% using --*mendel* in PLINK 1.9 beta (Chang et al., 2015, Purcell and Chang, 2017). Parental relationships were also tested using the program SEEKPARENTF90 (Aguilar, 2014); however, this program was not as effective if genealogically assigned broodstock had not been genotyped. Depending on the

generation, this resulted in 23-79% of the offspring not being tested for parentage. Of those assignments that could be conducted, reassignment rates were within the same range as those obtained in the above described pipeline.

When comparing molecular assignment results to pedigree records, each case was classified into one of four categories: pedigree agreement, reassigned sire or dam, indiscernible sire or dam, or unknown error status (Appendix 1). In pedigree agreements, both the molecular and pedigree-based assignments were in agreement. Reassigned sires or dams were those parents who could be correctly assigned using molecular data, but where the molecular and pedigree data did not agree. Indiscernible sires or dams occurred when an offspring's pedigree parents were not assigned to them using molecular data but could not be reassigned since their true parents were not genotyped. Unknown error status was given to those parents whose assignment could not be confirmed due to only partial reassignments of parents that could not be confirmed using the pipeline described above. Assignment mismatches were only identified as 'pedigree errors' (reassigned or indiscernible sires or dams) if both parents could be reassigned, or if both parents had been genotyped and one or more of the assigned pedigree parents was not genetically assigned to the offspring. If both parents were not genotyped, then any partial reassignments (i.e. only a single parent) were not classified as an error since having at least one unknown parental genotype could introduce ambiguity in relation to the source of the error. Hence, these cases were classified as having an unknown error status. If only one parent was genotyped then any partial reassignments (i.e. only a single parent) were not classified as an error, unless their pairwise genetic relationship confirmed that the parent was not the true parent.

To determine whether or not pedigree errors were family based (i.e. genetically assigned parents were siblings to the listed pedigree parent) relatedness was calculated using both molecular and

pedigree data for all offspring in generations 10 and 11 who could be assigned at least one parent in generations 9 and 10, respectively, using molecular data. Relationship matrices for pedigree data were built using TASSEL v. 5.0 (Bradbury et al., 2007) for pedigree data and GCTA v.1.26.0 for molecular data (Yang et al., 2011).

2.2.4 Pedigree Genetic Diversity

Calculations of observed heterozygosity (H_o), expected heterozygosity (H_e), and the number of monomorphic SNPs for generations 9 to 11 were conducted using a Markov Chain (MC) length of 1,000,000 with 100,000 dememorizations in ARLEQUIN v. 3.5 (Excoffier and Lischer, 2010). The average inbreeding coefficient (*F*) per generation, based on the observed and expected number of homozygous genotypes ($\hat{F}^{ii}$), was established using the command *–ibc* in GCTA v.1.26.0 (Yang et al., 2011). The average multilocus heterozygosity (MLH) for each population was computed using the R package inbreedR (Stoffel et al., 2016). Finally, the effective population size (N_e) per generation, based on the linkage disequilibrium method (LDN_e), was estimated in NeEstimator v.2.01 (Do et al., 2014).

2.2.5 Pedigree genetic structure

To understand the genetic structure of the AS, individual identity-by-state (IBS) distance matrices from generations 9 to 11 were first constructed using *–distance-matrix* in PLINK 1.9 beta (Chang et al., 2015, Purcell and Chang, 2017). A model-based estimation of ancestry was produced in Admixture v.1.3 which was then used to determine the number of family groups in our subsampling of the AS at a fivefold cross-validation rate (Alexander et al., 2009). Relationships among individuals comprised from the previously constructed distance matrices and admixture proportions were then visualized using NETVIEW v.1.1 at kNN 30 (Neuditschko et al., 2012, Steinig et al., 2016).

2.3 Results

2.3.1 Genealogical Analysis

While genealogical records were available for generations 1-11, tissue samples for molecular analysis were only available for parents (generations 9 and 10) and offspring (generations 10 and 11) of the AS. Subsequently, only those generations with molecular data could be corrected for pedigree errors. On average, 45.5% of AS pedigree records were erroneous per generation (Fig. 2.1) when comparing parents to offspring (i.e., G9 parents with G10 progeny & G10 parents with G11 progeny). Generation 11 exhibited the highest error rate (54.9%), but with the greatest percentage of its broodstock (generation 10) genotyped (81.2%) it also had the highest pedigree reassignment rate for identified errors (64.3%), resulting in 35.3% of records in that generation being corrected (Fig. 2.1). Generation 10, which only had 62.4% of broodstock (generation 9) genotyped, had a reassignment rate of 52.1% with 25.7% of the errors in that generation corrected (Fig. 2.1).

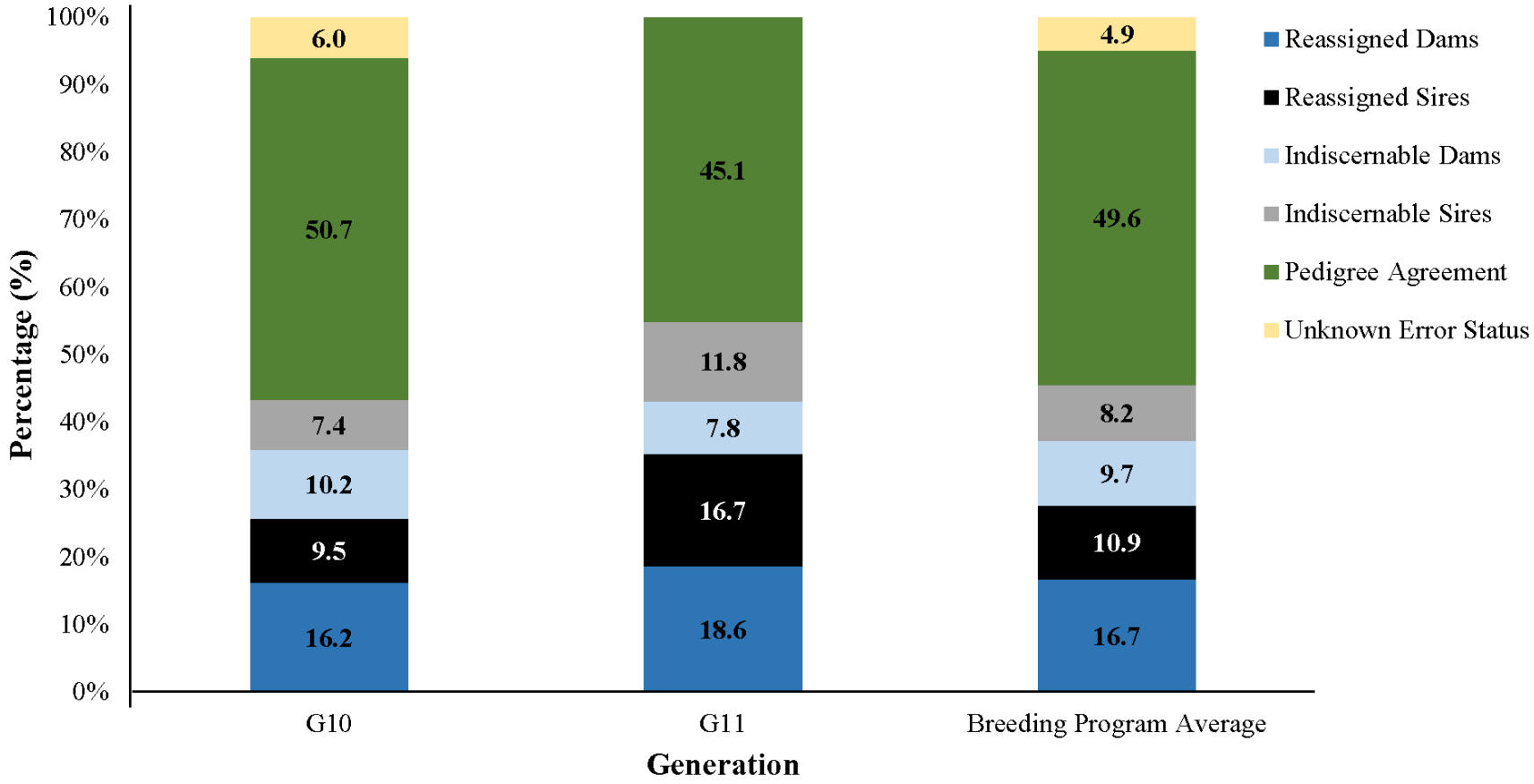


Fig. 2.1 Comparison of genealogical and molecular pedigrees. Genealogically and molecular derived pedigrees were compared and genealogical records were corrected according to molecular records. Records were categorized as 'Pedigree Agreement' (green), 'Unknown Error Status' (i.e. those individuals whose original parents had not been genotyped and who could not be reassigned both parents with at least 95% certainty; yellow), Reassigned Dams (i.e. those offspring who were reassigned a true dam, blue), Indiscernible Dams (i.e. offspring whose genealogically assigned dam was incorrect, but whose correct dam had not been genotyped and therefore could not be assigned, light blue), Reassigned Sires (i.e. those offspring who were reassigned a true sire, black), and Indiscernible Sires (i.e. offspring whose genealogically assigned sire was incorrect, but whose correct sire had not been genotyped and therefore could not be assigned, grey).

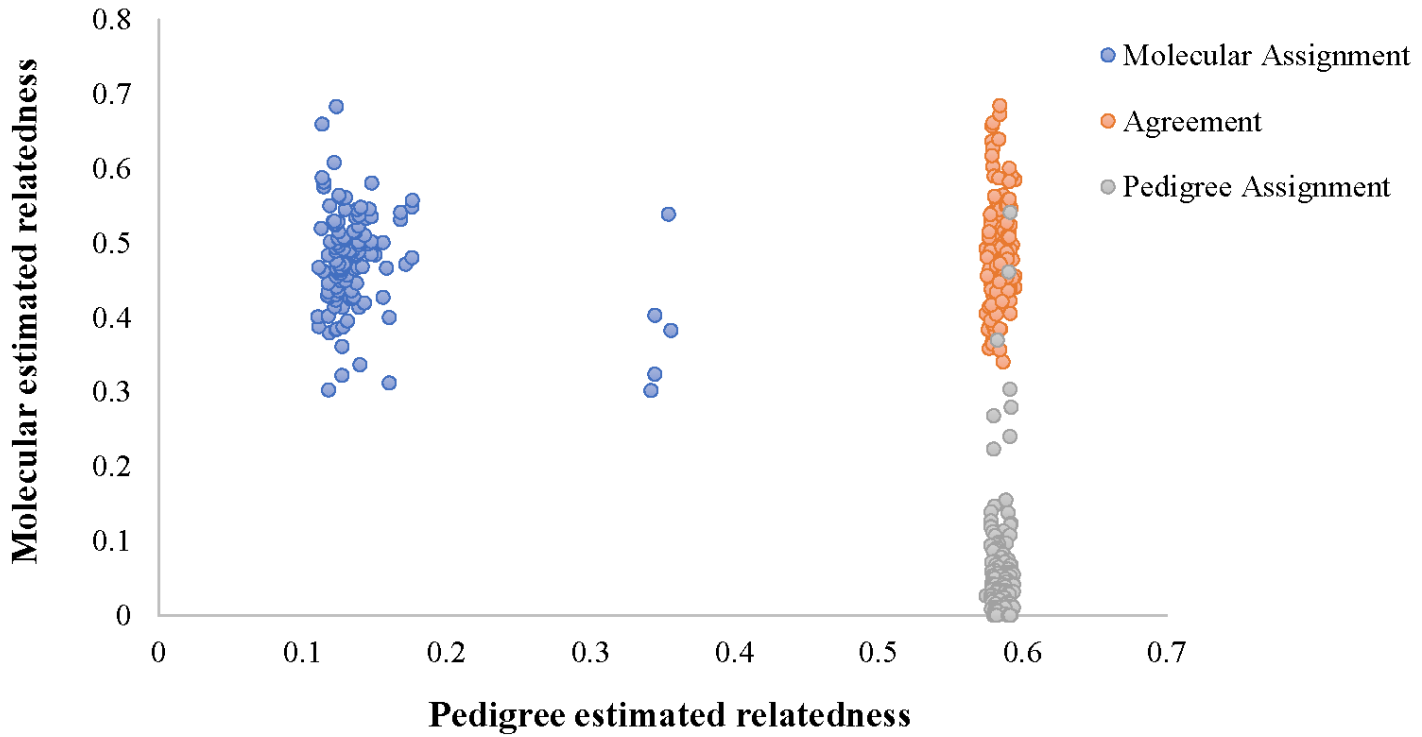


Fig. 2.2 Comparison of pairwise molecular and pedigree estimated relatedness for parents and offspring pairings. Colour designations reflect assignment and re-assignment classifications used during pedigree analysis. Comparisons of relatedness are between molecular and pedigree identified parents (whether via molecular assignment, pedigree assignment or agreement between both methods) as calculated using molecular data and their original pedigree relationships. Parents whose assignment agreed in both molecular and pedigree estimates are denoted in orange (Agreement). Molecularly assigned parents which did not agree with the pedigree data are denoted in blue (Molecular Assignment).These fell into two distinct groups ranging from 0.1-0.2 in the pedigree relatedness scale (indicating random errors) and 0.3-0.4 in the pedigree relatedness (indicating family-based errors) Those originally assigned pedigree parents not in agreement with molecular data are denoted in grey (Pedigree Assignment). Errors were categorized as random errors from 0-0.1 on the molecular relatedness scale and family-based errors from 0.1-0.5 in the molecular relatedness scale. Please note that molecular ranges vary due to biases resulting from missing data in genetic relationship calculations. Molecular based relationships were scaled per individual based on the relatedness calculated to self. These ranges were then confirmed by the number of Mendelian Inheritance errors detected per molecular family grouping (<3%).

A comparison of molecular- and pedigree-based relatedness estimates revealed three clusters that reflect the categories assigned during the investigation of pedigree relationships. As molecular relationships were sensitive to the proportion of missing SNP data, a range of relatedness between 0.30-0.70 was accepted as a parental match if all other conditions for reassignment were met. This range resulted in less distinct clustering; however, relative groupings were still distinguishable. The cluster with high molecular relatedness (0.32-0.55), but low pedigree relatedness (0.11-0.36) consists of individuals that had parentage reassigned through molecular analysis (Fig. 2.2). The cluster with high pedigree relatedness (0.57-0.59), but low molecular relatedness (0-0.54) consists of original pedigree assignments that differ from current molecular assignments. The cluster with both high molecular relatedness (0.34-0.69) and high pedigree relatedness (0.57-0.59) consists of those parental relationships that were in agreement between both pedigree and molecular assignments. Of those offspring who had been reassigned, 19.6% of incorrect pedigree assignments were reassigned to siblings of their originally assigned parents and 4.3% of molecularly assigned individuals corresponded to parental siblings based on pedigree data (Fig. 2.2).

2.3.2 Founder Contributions

Correction of genealogical records was dependent on the available molecular data, and subsequently could only be conducted for generations 10-11. Although genealogical data were shown to have higher error rates than expected, uncorrected genealogical records were used as a benchmark for expected levels of founder contributions, genetic diversity, and inbreeding within the AS. These were then compared to results from molecular data to quantify the effects of pedigree errors.

Despite variations between original and corrected genealogical records, calculations for the total number of founders (f = 73) and the effective number of founders (f^e = 46.6) were in agreement between both datasets for the most recent generation (G11). Pedigree data indicates that only 83 of the original 201 founders contributed to subsequent generations (Appendix 2). Of these 83 founders, the family lines of only 53 founders (28 dams and 25 sires) could be traced to corrected generation 11 records with a minimum contribution of 0.1% (Fig. 2.3). Similar results were also observed in estimates of effective population size (LDN_e) for generations 9 (LDN_e = 52.6-53.7) and 10 (LDN_e = 53.9-54.7). Only the LDN_e data of generations 9 and 10 were reported as over 60% of available broodstock were sampled in these generations, whereas a sampling bias occurred in generation 11, with less than 20% of broodstock sampled in this generation. Of the 53 founders, 47 have a traceable contribution of greater than 0.3% to generation 11, which corresponds to the effective number of founders of 46.6 (Fig. 2.3; Appendix 3).

During the 4th generation, 2,178 individuals from another domesticated line were introduced into the AS program, but only 94 individuals were utilized as broodstock for generation 5 of the AS (Appendix 2). Of these 94 broodstock, only 46 individuals had a traceable genomic contribution of 0.1% or greater to generation 11 of the AS (Fig. 2.3). Of these 46 secondary founders, all but one individual (1.5%) contributed less than 0.3% to generation 11 genomes. The greatest loss of genetic material from the original 83 founders (14 individuals, 16.9%) occurred in generation 5 after the secondary founders had been incorporated in the AS (Appendix 2). The number of original founders whose genetic contribution could be traced continued to drop throughout generations 5-9, with a total of 30 (36.1%) founder genomes lost (Appendix 2). Founder contribution remained consistent throughout generations 9-11. Original founders accounted for 92.0% of AS genetic material (with 34 original founders comprising 84.3% of the AS's genetic

composition) and generation 4 secondary founders comprising 7.3% (Fig. 2.3). Across corrected generations (G9-11), only 0.7% of the founder contribution remained undetermined due to unassigned pedigrees in corrected genealogical files (Fig. 2.3).

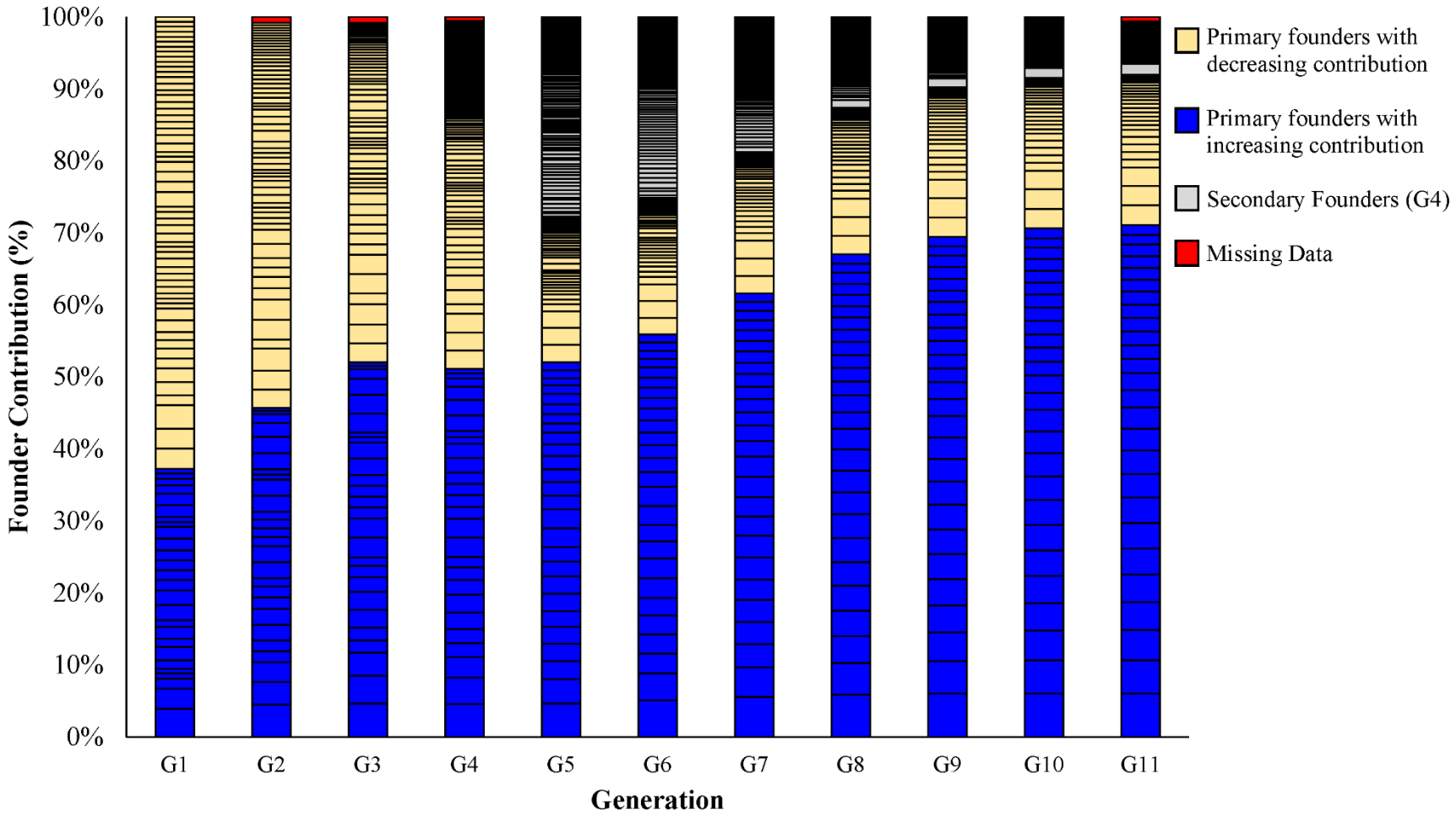


Fig. 2.3 Genome contributions of founders over 11 generations based on corrected pedigree records. Contributions are based on each founder's average contribution to each generation's genome pool. Contributions have varied over the generations with original founders whose contribution increased in subsequent generations (blue), original founders whose contribution diminished or is negligible over time (tan), contribution of individuals introduced in generation 4 (grey), or unknown contribution due to missing pedigree records (not recorded or due to indiscernible pedigree errors, red). Due to border edge effects, founders with minimal contribution may be indistinguishable from one another (black).

1

2.3.3 Genetic Diversity Indices

Molecular data reveals that while both H_e and H_o remained relatively constant from generations 9-11 and at a ratio of approximately 1:1 (H_o/H_e; Table 2.1). *F* increased by 0.036 between generations 9 and 10, (F_{G9} = 0.044*;* F_{G10} = 0.080*;* Table 2.1). This was followed by a decrease in *F* in G11 to 0.041; however, this is likely due to sampling bias as only 19.0% of G11 was sampled (Table 2.1). This was accompanied by an increase in monomorphic SNPs (207) and a decrease in multi-locus heterozygosity between G9 to G11 (3.7%; Table 2.1).

Table 2.4 Molecular measures of genetic diversity. Expected heterozygosity ($H_e \pm$ SD), observed heterozygosity ($H_o \pm$ SD), multi-locus heterozygosity (MLH± SD), inbreeding coefficient ($F \pm$ SD), and monomorphic SNPs are presented for generations 9-11 using molecular data.

	G9	G10	G11
H_e	0.225 ± 0.144	0.223± 0.146	0.225 ±0.147
H_o	0.226 ± 0.153	0.221 ± 0.153	0.221 ± 0.157
MLH	0.215 ± 0.025	0.204 ± 0.031	0.207 ± 0.019
F	0.044 ± 0.063	0.080 ± 0.074	0.047 ± 0.020
Monomorphic SNPs	20	15	227

Based on the original genealogical data, there is a steady increase in inbreeding from founders to generation 11 (ΔF = 0.003-0.019 per generation) with a slight decrease (ΔF = -0.010 in inbreeding in generation 5 due to the new individuals introduced in generation 4; Fig. 2.4). The average increase in inbreeding per generation in corrected records was ΔF = 0.004. Corrected pedigree records exhibited a slight decrease in inbreeding in the corrected generations and generation 4. It should be noted that these corrected records include a higher number of records with one or more missing parents in generations 10 and 11, with PEDIG assuming unknown parents to be new and unrelated individuals (Boichard, 2002). This may result in inbreeding coefficients being underestimated in these generations. Corrected and original inbreeding coefficient estimates based on genealogical records for generation 10 ($F_{Original}$ = 0.057; $F_{Corrected}$ = 0.056) were approximately 25% less than those values estimated using molecular data (G10, F = 0.080). The variation between original and corrected pedigrees in generation 4 can be attributed to the original exclusion of founder information for those introduced animals (Fig. 2.4)

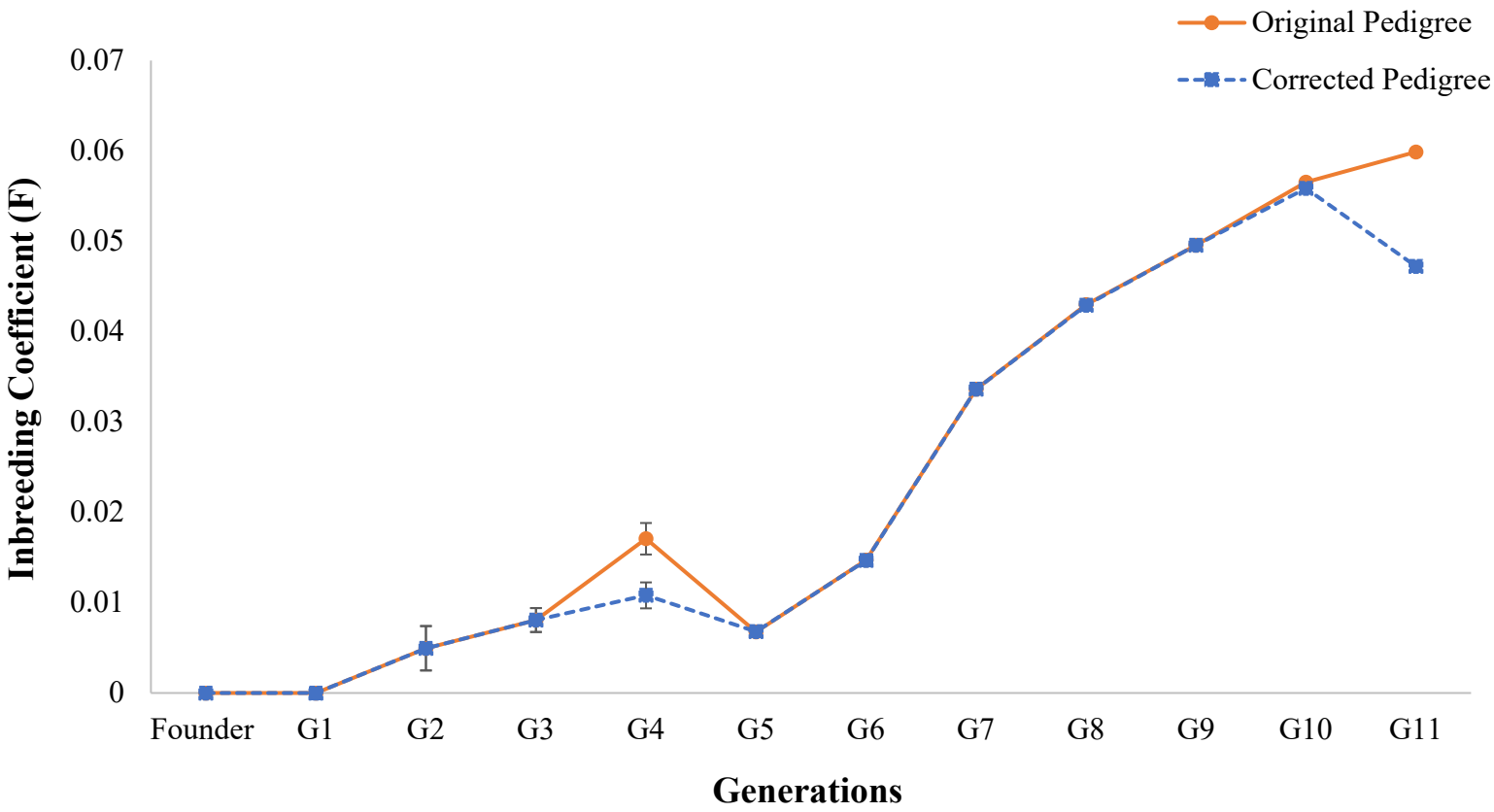


Fig. 2.4 Inbreeding coefficient (*F*) for original and corrected genealogical records. Inbreeding based on genealogical records was calculated using the method described by Luo (1992), which utilizes the Cholesky factor of the relationship matrix. Inbreeding coefficients were calculated for both the original pedigree assignments (orange, solid line) and pedigree records corrected using molecular data (blue, dashed line). Error bars represent the standard error of the mean (SE) for each generation. Drops in molecular inbreeding coefficients coincide with the addition of previously undocumented new germplasm (G4) and a bias in the PEDIG program in which individuals with unknown parents are treated as new individuals, underestimating the true level of inbreeding within the program (G10-G11).

2.3.4 Pedigree Genetic Structure

A total of 27 family lines were identified throughout generations 9-11 using Admixture v. 1.3 and NETVIEW v.1.1 (Fig. 2.5). As only 53.3% of available broodstock were genotyped this number is likely underestimated for the total AS genetic stock. Each node, or individual depicted as a circle, illustrates the relatedness of that individual to all other individuals sampled. Colours within the node indicate admixture, with the AS displaying high admixture amongst families (Fig. 2.5). The larger the node, the greater the proportion of the population to which that individual is related. Node size of individuals from the oldest available generation (G9) were all approximately equal in size, suggesting that families contributed equally to subsequent generations (Fig. 2.5). The thickness of the lines, or edges, connecting individuals represents the genetic distance amongst individuals. In this case, the lines exhibit similar weights, indicating a similar level of genetic distance amongst individuals. Similar edge weights in addition to short distances and high connectivity between individuals suggests that there is a high degree of relatedness within the AS (Fig. 2.5).

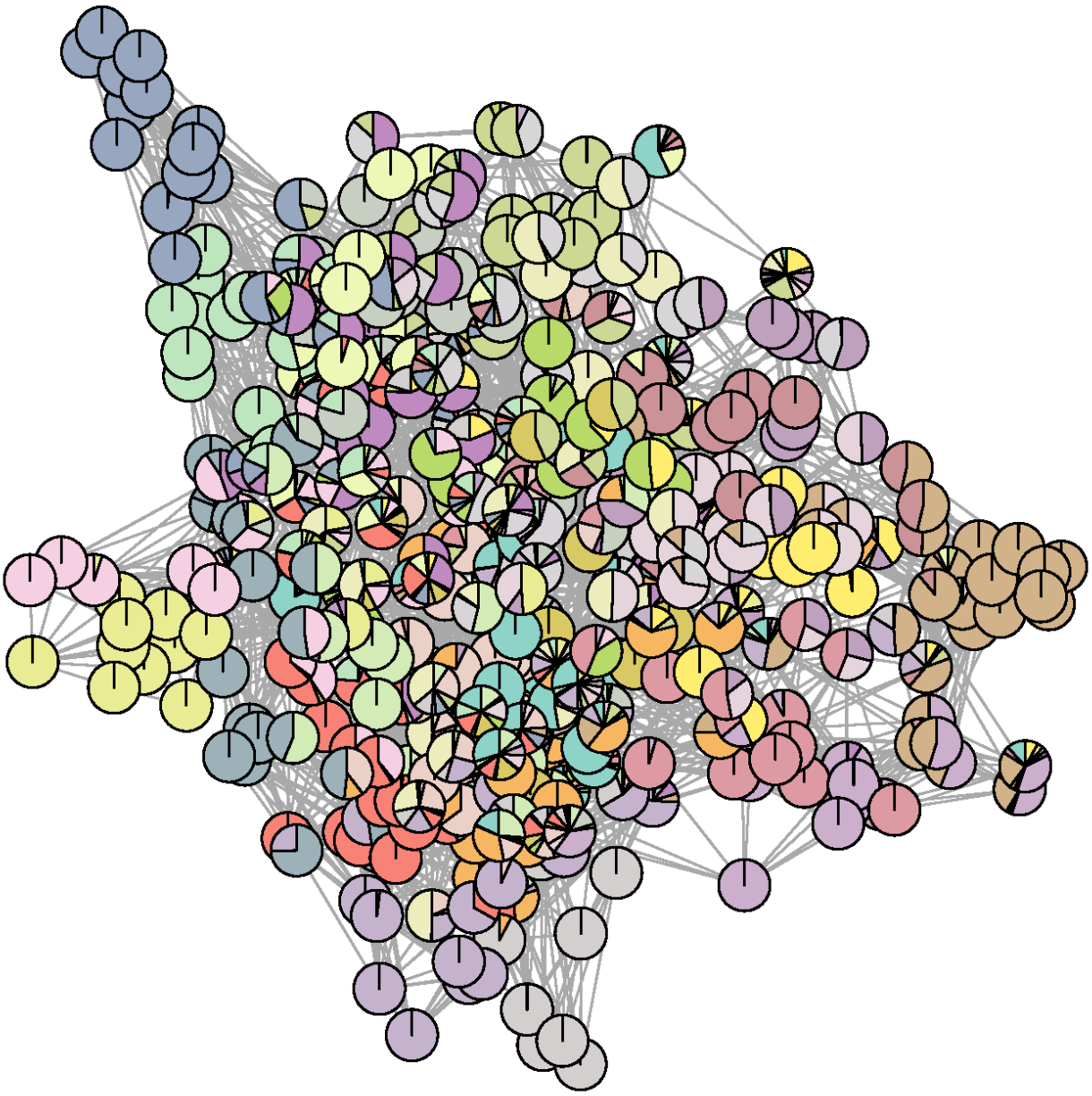


Fig. 2.5 High-resolution population structure of generations 9-11 of the Abbassa Strain (AS). Network visualization of 388 Nile tilapia from three generations of the AS where individuals are represented by a node. Node colour(s) represent individual family lines, with solid colours representing family founders, node size is indicative of relatedness, node pie charts exhibit ancestry proportions, linking lines (edges) representing genetic similarity, and edge weight reflecting genetic distance. The patterns show that individuals from Generation 9 have all contributed approximately equally to subsequent generations (i.e. node sizes are similar). Twenty-seven individual family lines were identified via cross-validation plots. The AS displays high admixture amongst families.

2.4 Discussion

This study used genome-wide molecular markers to 1) determine the factors that contributed to the relatively slow rate of improvement observed within the AS, 2) understand the current state of the strain and 3) assist in determining future strategies for improving farm management.

One key issue which has likely attributed to the slow rate of improvement observed within the AS were the high levels of pedigree errors identified within the study. An average total error rate (including both reassigned and indiscernible parents) of 45.5% was observed in pedigree records. At over three times higher than the maximum observed pedigree error rate (1-15%) in other selective breeding programs, this is a significant difference between the AS and other systems (Bovenhuis and Van Arendonk, 1991, Crawford et al., 1993, Sanders et al., 2006). Discussions with the breeding team revealed a number of issues that occurred over time: such as, multiple naming protocols, unclear IDs, transcribing errors, historical errors (i.e. errors from previous generations causing downstream misidentification) and changes in management. In addition to unforeseen mixing of untagged fish amongst hapas (fine mesh net enclosures) due to the close proximity of hapas and during flooding events. All of these factors could contribute to an accumulated negative effect on the breeding program.

Once detected, 52.1-64.3% of pedigree errors were reassigned to their correct parent pair within this study. These reassignment rates could be improved by genotyping more broodstock from within respective generations so that family groupings with unconfirmed parents can be assigned their true parentage. As pedigree errors remained undetected until molecular data was available, it is suggested that genotyping be incorporated into selective breeding programs as early as possible to not only correct, but to prevent the accumulation of these errors.

With weight estimated to have a heritability of 0.15-0.41 for *O. niloticus* (Khaw et al., 2009), the lower genetic gain observed within the AS of *O. niloticus* (3.8-7.0%; Rezk et al., 2009) compared to the GIFT (Genetically Improved Farmed Tilapia) strain (7.1-15.0%; Eknath and Acosta, 1998, Ponzoni et al., 2011) could also have resulted from a range of factors including differences in environment and the base genetic content of those two populations. However, given the high pedigree error rates, it is quite likely that the management of the AS has attributed to a reduction in selective pressure on the AS. Of these errors, pedigree errors that were family based (19.6%; i.e. offspring assigned to siblings of original pedigree parents) likely contributed less to this reduction in genetic gain than other error types.

Despite these high error rates, the AS's current genetic diversity after 11 generations is still sufficient for the continuation of the breeding program. The AS has on average remained under the 1% increase in inbreeding coefficient per generation deemed acceptable for selective breeding programs (Woolliams, 1994, Bentsen and Olesen, 2002). Additionally, molecular diversity indices do not show a significant drop in observed, expected, or multi-locus heterozygosity amongst generations nine, 10, or 11. The inbreeding coefficient (*F*) calculated using molecular data for G9 and G11 are similar to the inbreeding coefficients calculated using original and corrected pedigree data; however, molecularly calculated *F* is 25% greater than pedigree estimates of *F*. This suggests that pedigree records may provide acceptable estimates for inbreeding rates within the AS; however, these may be an underestimate of the true level of inbreeding due to missing data from individuals with unknown parentage. This is a constraint of PEDIG, which views unknown relationships as new genetic material, the estimates of inbreeding rates are considered to represent the lower bounds of what is observed within the farm stocks (Boichard, 2002).

In order to maintain adequate genetic diversity for genetic improvement and adaptability during dissemination, aquaculture management schemes focus on the initial number of founders in their selective breeding lines and subsequently the number of families retained per generation (Skaarud et al., 2014). However, our analyses have shown that maintaining family lines does not ensure that founder genomes are retained. Of the initial 201 founders only 83 contributed to generation 1, with both pedigree and molecular data indicating that only 53 founder genomes were still present in generation 11. While only an approximation due to pedigree errors, pedigree data revealed that of these 53 founders, only 34 (41.0% of original founders) comprised approximately 84.3% of the AS's genetic composition despite maintaining approximately 110 full-sibling families per generation for this line. Considering the majority of the founders for the AS came from wild populations, unequal founder contribution is somewhat expected due to failed spawning, batch drop out, domestication selection, or undesirable trait combinations from untested stock. However, this high loss of genetic material through unequal contributions of founders during generation 1 and generations five to nine of the AS program highlights the importance of preserving unique founder genomes throughout the program in order to obtain the highest level of genetic diversity possible within a closed system under selection. If family numbers rather than founder genomes are monitored, then only those individuals within a family that exhibit the desired traits most strongly will be selected as future broodstock. This encourages genetic gain, but limits genetic diversity within a program as these traits can be traced to a subset of high performing founders. As time progresses, the contribution of these high performing founder genomes continues to increase, as demonstrated in the present study. Consequently, it is suggested that alongside family lines, founder contributions also need to be monitored to avoid rapid loss of genetic diversity when managing selective breeding programs.

This should ensure the genetic health and longevity of a selected line in addition to maximizing the adaptation potential of the line in the future.

Cross breeding different domesticated lines has been shown to reduce the effects of inbreeding and increase genetic diversity within a selective breeding line (Goyard et al., 2008, Stronen et al., 2017); however, the legacy of these crosses in relation to the maintenance of genetic diversity over time is uncertain. In 2006, 2,178 individuals from another Abbassa mass selection breeding program, whose relation to the AS is unclear, were introduced into the AS, resulting in a 0.004 (ΔF) reduction in inbreeding in generation 4 of the AS based on corrected pedigree records. However, this reduction in inbreeding was short-lived since only a few of these individuals were selected as broodstock for subsequent generations and their unique genetic contribution diminished substantially within four years. A number of management considerations could have prevented the loss of these introduced genomes which would have not only reduced inbreeding but may have resulted in further performance improvement after having undergone the same selective pressures as AS animals. For example, the identification of novel founder lines throughout subsequent generations would have allowed selective practices to maintain their germplasm within the line, as well as allowing for selection of high performing individuals. The incorporation of this technique would improve the genetic gain over generations in addition to increasing the retention of germplasm, and therefore minimizing inbreeding and maximizing genetic diversity.

Accurate genealogical records are essential in optimizing the genetic gain obtained within a selective breeding program. Any inaccuracies in pedigree information will erode the accuracy of EBVs and diminish the rate of genetic improvement within the line (Banos et al., 2001, Israel and Weller, 2000). If recorded pedigrees are not available, or in question, molecular based

parentage analyses provide a practical and efficient means to identify parent and offspring relationships within a selective breeding program. Here once the pedigree of the fish has been ascertained through DNA parentage, marking fish with a physical tag will help maintain identity of the fish from this point.

However, there is a trade-off between obtaining false positives and negatives in assignments. Presently, there are three main categories of pedigree reconstruction methods: exclusion methods, relatedness-based methods, and likelihood-based methods (Huisman, 2017). Of these, likelihood-based methods are more powerful in assignment capabilities, although they require more computation time, particularly as the number of markers analyzed increases (Hill et al., 2008). Whilst both CERVUS and COLONY utilize a likelihood-based method, there are known limitations to these programs. To reduce computational time, CERVUS only considers pairwise likelihoods to find the most likely parent. However, this can result in close relatives who are not true parent and offspring pairs having a positive log-likelihood ratio, potentially resulting in a false positive (Huisman, 2017). This can be particularly problematic when not all parents have been genotyped. COLONY parental assignments are most accurate when analyses are conducted with highly polymorphic markers; if only less informative markers are available, higher rates of incorrectly assigned parents are observed (Jones et al., 2010). This can be compensated for by using a higher number of markers, with error rates becoming negligible in COLONY ($<4.5E^{-4}$) when 75 or more SNPs are used for analysis with 40% of parents not genotyped (Huisman, 2017). As such, we found that an integrated workflow including the likelihood-based methods described above, pairwise relatedness estimates and exclusion-based (Mendelian inheritance errors) methods should be utilized to ensure that only reliable parentage relationships are produced.

2.5 Conclusions and Industry Recommendations

The prevalence of pedigree error rates in AS were found to be three times greater than in terrestrial programs (45.5% on average). This high pedigree error rate is likely to have contributed to the low levels of genetic gain (3.8-7.0%) per generation observed within the AS, but did not appear to have a major effect on overall inbreeding levels. While the AS aims to produce 100-120 families per generation, only 28 unique family lines were identified based on genotyping 57.1-63.9% of available broodstock from generations 9-11. An assessment of founder contribution to the AS revealed that only 34 founders comprise over 84.3% of available genetic material within the AS, indicating that founder contribution has been eroded within the AS. These results suggest a review of the breeding program procedures is necessary, in particular a tightening of the family production process is required to reduce the likelihood of errors; however, the overall genetic composition of the AS population is sound and still provides an acceptable basis for continued breeding. The results also suggest that the inclusion of molecular screening should be introduced, preferably at program inception, to ensure greater pedigree accuracy. It is also recommended that when dealing with closed populations, particularly selective breeding programs, and the retention of founder genomes throughout the program should be monitored in addition to family lines.

CHAPTER 3: NOVEL POPULATION BASED LINKAGE MAPPING, QTL AND GWAS FOR SEX AND GROWTH WITHIN THE ABBASSA STRAIN OF NILE TILAPIA (*Oreochromis niloticus*)

3.1 Introduction

To improve the efficiency of fish production, the aquaculture industry has turned towards selective breeding. Selective breeding uses the natural genetic variation within a founding population to breed subsequent generations of animals with a higher frequency of desired marketable traits. In 2002, the Abbassa Strain (AS) of Nile tilapia (*O. niloticus*) was established by the WorldFish Center in Egypt with the objective to increase weight using a combination of between and within family selection (Ibrahim et al., 2013, Rezk et al., 2009).

At present, the AS selective breeding program for Nile tilapia has been based on phenotypic information alone. While this approach has been successful and has resulted in a realized 3.8-7.0% improvement in growth per generation (Rezk et al., 2009), productivity could be further improved through marker-assisted selection (MAS) or genomic selection (GS). While similar, these two methods mainly vary in how markers are used to estimate breeding values, with MAS based on quantitative trait loci (QTLs) and working best with genes of major effect (Arruda et al., 2016), and GS using all available high-quality markers, accounting for all QTLs associated with a trait, regardless of their effect size (Goddard and Hayes, 2007).

Both MAS and GS rely on the concept of "hitchhiking," in which the regions surrounding a gene being selected for are also selected, leaving signatures of selection in areas adjacent to the gene of interest (Smith and Haigh, 1974). To identify these signatures of selection in the AS, it is first

useful to create a robust, strain specific, high-density genetic linkage map. In turn, this map can be used to understand the effects of selective breeding on the *O. niloticus* genome and the genetic architecture of specific traits, like sex or weight, through QTL and GWAS analyses (Du et al., 2016, Tsai et al., 2015). If a gene of major effect is detected, it can then be used to direct MAS; alternatively, if the trait is polygenic, all identified QTLs can be incorporated into a GS statistical model to improve prediction accuracy (Zenger et al., 2019).

To date, there are two genome assemblies (*O. niloticus* Orenil 1.1 and O_niloticus_UMD1) and five linkage maps published for *O. niloticus* (Table 1.1; Conte et al., 2017, Guyon et al., 2012, Joshi et al., 2018, Kocher et al., 1998, Lee et al., 2005, NCBI, 2017, Palaiokostas et al., 2013), although, neither is specific to the AS. Genome assemblies and linkage maps are two complimentary genomic resources in which linkage maps identify the relative order of markers to one another on a chromosome, whereas genome assemblies, or physical maps, give the physical distances of markers to one another. Both available physical maps have been assembled to the chromosomal level; however, they both still have a substantial number of unplaced scaffolds (2,460-5,655; NCBI, 2017). These unplaced scaffolds can be problematic for molecular studies as genes of interest may be fractured or incorrectly annotated (Baker, 2012, Denton et al., 2014). Genetic linkage maps can be utilized to improve physical genome maps (Fierst, 2015). These maps rely on meiotic recombination rates to determine the relative position of markers, and these recombination rates can vary greatly between species, populations, individuals and even genomic regions (Dukić et al., 2016). As such, it is imperative to create line specific high-density genetic linkage maps for selectively breeding programs, like the AS for Nile tilapia.

Once a high-density genetic linkage map is created for the AS, the traits of sex and weight can be further explored. Sex determination in *O. niloticus* is largely genetic, with a male heterogametic

(XX|XY) sex determining system (Mair et al., 1991); granted, environmental factors have also been proven to affect sex ratios and cause sex reversal in fry (Baroiller et al., 2009, Wessels et al., 2014). To date, sex related markers have been associated with LG 1 (Conte et al., 2017, Palaiokostas et al., 2015), LG 3 (Palaiokostas et al., 2015), LG 8 (Lee et al., 2003), and LG 23 (Cáceres et al., 2019, Eshel et al., 2011, Palaiokostas et al., 2015). There have been relatively few QTL studies in Nile tilapia for weight or growth rate, with most based on interspecific hybrids (Lin et al., 2016, Liu et al., 2014). Weight related markers have been detected in LG 1, LG 3, LG 7, LG 10 (Liu et al., 2014), LG 12 (Lin et al., 2016), LG 13, LG 19 (Liu et al., 2014) , LG 20, and LG 22 (Lin et al., 2016) in Nile tilapia and other tilapia species including, *O. aureus* which readily hybridize with Nile tilapia (D'Amato et al., 2007, Deines et al., 2014, Lovshin 1982, Meier et al., 2019). Despite numerous studies, the polygenic nature and the influence of gene by environment (GXE) on both sex and weight make trait architecture difficult to unravel and there is still a great deal that we do not understand about both of these complex traits (Baroiller et al., 2009, Cáceres et al., 2019, Conte et al., 2017, Eshel et al., 2011, Lee et al., 2003, Liu et al., 2014, Mank, 2008, Palaiokostas et al., 2015, Wang et al., 2019); particularly, how and if relevant QTLs vary amongst different strains of Nile tilapia.

This chapter aims to investigate the genomic trait architecture of weight and sex in the Abbassa Strain of Nile tilapia by 1) constructing an independent framework genetic linkage map for the AS, 2) comparing and evaluating the AS framework map against genome sequence assemblies, 3) understanding the sex determination system (chromosomes vs. regions) in AS tilapia, and 4) resolving the genetic architecture of growth traits in relation to weight.

3.2 Methods

3.2.1 Phenotypic Data

Phenotypic data from 388 individuals [122 animals from generation 9 (G9); 216 animals from generation 10 (G10); and 54 animals from generation 11 (G11)] were collected for two major traits: sex and weight. Animals were harvested at between 9-11 months of age, and sex and final weight recorded. In order to ensure that associations corresponded as strongly to phenotypic data as possible, information that could affect associations (earthen pond in which the fish were raised, age, and generation), treated as covariates, were also provided by WorldFish. Raw phenotypic data for final weight at harvest was adjusted for age and analyzed for family and individual differences.

3.2.2 Reference Mapping Families

A mapping resource was identified within the sample set consisting of 16 families ranging from 5-17 offspring per family as established in Chapter 2 (Table 3.1). Across these families, a total of 166 individuals were used to provide sufficient resolution for linkage mapping analysis across thousands of genetic markers with 18 individuals used in QTL analysis (Table 3.1). A total of 388 individuals (including the 166 individuals belonging to families used in mapping) from across Generations 9-11 were used in GWAS. There was some overlap in individuals used as offspring and parents for the two generational family groupings; in total, these families consisted of 32 F_0 and 136 F_1 offspring. Eight of these families (52 individuals) were founded from generation 10 of the AS, while the remaining eight families (84 individuals) were founded from generation 11. To create more robust families for linkage mapping purposes, families in generation 11 included 76 samples which were harvested when only fingerlings. As phenotypic data was not collected on farm for these samples, they were excluded from genome-wide association studies (GWAS) and quantitative trait locus (QTL) analyses.

To determine the detection power in this study, the GCTA-GREML Power Calculator (Visscher et al., 2014) was used based on different sample sizes, heritability estimates (0.38-0.60) for growth-rate were based on Charo-Karisa et al. (2006) and type 1 error rates, and variance of SNP-derived genetic relationship based on Yang et al. (2011).

Table 3.5 Summary of families and their use in analyses.

Family	Dam	Sire	Total Offspring	Fingerling Offspring	SilicoDArT Markers	Linkage Mapping	QTLs	GWAS
1	9003575	9004941	5	0	N	Y	N	N
2	9005019	9003328	5	0	N	Y	N	N
3	10000809	10002618	5	5	N	Y	N	N
4	9000315	9001183	6	0	N	Y	N	N
5	9001405	9002879	6	0	N	Y	N	N
6	9004241	9005806	6	0	N	Y	N	N
7	9005343	9002914	6	0	N	Y	N	N
8	10000983	10004046	6	6	N	Y	N	N
9	10002408	10003982	6	2	N	Y	N	Y
10	9000136	9002112	8	0	N	Y	Y	N
11	10003956	10002076	9	6	N	Y	N	Y
12	9004197	9004306	10	0	Y	Y	Y	N
13	10000991	10002136	12	11	Y	Y	N	Y
14	10005644	10002203	13	13	Y	Y	N	N
15	10001974	10005497	16	16	Y	Y	N	N
16	10004724	10002267	17	17	Y	Y	N	N

3.2.3 Sampling, DNA Extraction, Genotyping, and SNP Filtering

Fin clips from 486 samples were collected from generations 9-11 of the AS [121 individuals from generation 9 (G9); 216 individuals from generation 10 (G10); and 146 individuals from generation 11 (G11)] of the Abbassa Strain (AS). DNA extractions and genotyping were conducted by Diversity Arrays Technology as described in Lind et al. (2017) and in section 2.2.2.1.

To ensure that only high-quality and informative data was maintained for subsequent analysis, individuals with greater than 30% missing data were first removed and the returned genotypic data was filtered as described in section 2.2.2.2. Pedigree errors were then corrected as described in section 2.2.3. Subsequently, SNPs were filtered for Mendelian Inheritance (MI) errors using PLINK 1.9 beta (Chang et al., 2015, Purcell and Chang, 2017).

SilicoDArT markers, or genetically dominant markers whose data indicate the presence or absence of sequence variants, were also utilized for the families with 10 or greater offspring (5 families). The markers were then stringently filtered for a call rate greater than 95%; a one ratio greater than 0.001; a polymorphism information context (PIC) greater than 0.002; a read count greater than 8; reproducibility greater than 0.99, and a QPMR between 1.5 and 100 to ensure only SilicoDArT markers of the highest quality were retained.

3.2.4 Map Construction and Genome Coverage

Statistically, when using two generational families to create genetic linkage maps, moderate to large reference families (more than 50 progeny) are recommended to provide sufficient informative meiotic events to form linkage groups and separate closely spaced markers (Liu, 2017). Smaller families, with fewer informative meiotic events, result in lower pairwise LOD power for linkage statistics (Flaquer and Strauch, 2012). The number of informative meiotic events can be bolstered with a higher density of markers (Littrell et al., 2018); however, separation of these markers from common bins may be difficult. In addition, it is also important to have multiple mapping families since a single pair of parents may be homozygous at a locus of interest, rendering it uninformative in a single family (Liu, 2017). Due to the high pedigree error rates identified after sampling, only small families of 5-17 offspring were identified (Nayfa et al., 2020; Chapter 2). To address this limitation, population-based linkage mapping based on pairwise recombination data was utilized rather than more traditional family-based methods (Stam, 1993, Van Ooijen, 2018). Although this method bolstered mapping power, there was a constraint in mapping ability, with unique maps needing to be constructed for sex average, female and male lines, making direct comparisons among maps difficult.

To optimize the number of markers that could be mapped per linkage group, three different *de novo* mapping methods were compared using the three largest mapping families: family-based maximum likelihood (ML) mapping, family-based regression mapping based, and population-based regression mapping in JOINMAP v. 5 (Van Ooijen, 2018). As a population-based ML method was not statistically viable, this method was excluded for comparison. There were 10.2% more markers placed using family-based ML mapping than regression, but 24.8% more markers placed using population-based regression mapping than family-based regression mapping (16.2%

more markers than family-based ML mapping; data not shown). This variation in method performance is likely attributed to the small mapping families comprised of phase unknown data (i.e. only two generations of information available) available for this study. Mapping power was bolstered by combining all family information into a pairwise data population. Thus, population-based regression mapping in JOINMAP v. 5 was the optimal mapping method for this dataset (Van Ooijen, 2018).

Recombination frequencies and Z-scores per family following both maternal and paternal lines for linkage mapping were first calculated in LINKMFEX version 3.1 (Danzmann, 2016). Individual family recombination frequencies and Z-scores were then compiled together to create a pairwise data population. Pairwise data populations were created for both maternal and paternal lines, with these data combined to create sex average datasets.

To produce the most accurate genetic linkage maps, JOINMAP v. 5 was then used to first create *de novo* maps for all three datasets (female, male, sex average). Loci groupings were based on test for independence with a minimum LOD score of 4. As a population-based method was utilized, linkage maps from these groupings were constructed using regression mapping with the following thresholds: recombination frequency <0.4; LOD score > 1.0; and a goodness-of-fit jump threshold of 5. Finally, a ripple, or a moving window which considers all order permutations of three adjacent markers before selecting the best option, was conducted after the addition of each locus.

In addition to *de novo* mapping, SNP and SilicoDArT markers were also mapped using the established order of the two previously published genomes. To do this, SNP and SilicoDArT sequences were annotated to the two previously published genomes (Orenil1.1 and O_niloticus_UMD_NMBU) via BLAST2GO v. 5.2 (Götz et al., 2008). Markers that were

matched to a genomic position within either genome assembly were then assigned a starting marker order based on its linkage group and position in the respective genome assemblies. This order was then utilized as a fixed starting order within mapping analysis. As these maps were based on already established physical positions, mapping requirements were relaxed to a recombination frequency <0.5 and LOD score >0 while all other settings remained the same. While lower LOD scores were allowed, the linear regression method uses a weighted least squares procedure in which LOD scores are used as weights, therefore putting more weight on more informative data, i.e. those with higher LOD scores (Van Ooijen, 2018).

Once linkage groups were created, marker placement was confirmed using the nearest neighbor fit (N.N. Fit), a measurement used to indicate if markers are placed in a likely position (Van Ooijen, 2018). Any markers that were found to violate N.N. Fit (values ≥ 800cM) were removed. *De novo* maps were used to confirm the groupings and maps based on genome assemblies. If *de novo* linkage groups corresponded to two or more genome assembly groupings, *de novo* groupings were conducted at a higher LOD threshold and groupings compared again. In some instances, aberrant markers were detected (i.e. ≤2 markers from a *de novo* grouping would map to a second linkage group in a genome assembly grouping even at a higher LOD score) and those were removed.

In order to understand any discrepancies in mapping between both genome assemblies used, a synteny graph comparing all linkage groups in both assemblies was created using SynMap within CoGe (Haug-Baltzell et al., 2017, Lyons et al., 2008). Linkage maps for sex average, female, and male map groupings were visualized using the R package LinkageMapView (Ouellette et al., 2017).

3.2.5 Segregation Distortion

In order to identify markers which deviated from Mendelian Inheritance (MI), segregation distortion was examined using log-likelihood ratio tests for goodness of fit to MI expectations in LINKMFEX version 3.1. (Danzmann, 2016). In order to identify any sex specific or family specific segregation distortion, G-values were calculated for all markers in each family following both the maternal and paternal lines. G-values were then adjusted using a Bonferroni correction (mean family corrected alpha of 0.001).

3.2.6 Map Metrics and Inter-chromosomal Analysis

To identify sex specific differences amongst maps, metrics (including, total number of markers, number of unique markers, total map length, inter-marker distances, and marker density) and an inter-chromosomal analysis of marker order and position were conducted in the online platform *The Genetic Map Comparator* (David et al., 2017).

3.2.7 Quantitative Trait Locus (QTL) Mapping

Linkage mapping families were evaluated for their potential in QTL mapping within MapQTL v.6 (Van Ooijen, 2009). Only those families which could be phased in JOINMAP v. 5 (Van Ooijen, 2018) and which had phenotypic data recorded for traits of interest (sex and final harvest weight) were included in the analysis. These essential criteria reduced the dataset to only two families from Generation 10 with eight and ten offspring each. As only 18 individuals were viable for QTL analysis, the analytical power for QTL detection decreased and extra attention was taken in selecting the most appropriate approach to QTL mapping. MapQTL v. 6 offers three mapping methods to detect QTLs: nonparametric mapping (or Kruskal-Wallis analysis; KW), interval mapping (IM), and multiple QTL mapping (MQM). Briefly, KW is a

nonparametric equivalent of a one-way analysis of variance, IM utilizes either a maximum likelihood mixture model or regression mapping approach, and MQM is equivalent to a composite interval mapping approach and can also use either a maximum likelihood mixture model or regression mapping approach (Van Ooijen, 2009). MQM was initially selected as it builds upon the IM method and provides higher statistical power for analysis, and therefore sensitivity to the presence of multiple QTLs (Balding et al., 2007, Van Ooijen, 2009). However, due to small family sizes, families consisting of only two generations, and limited informative markers within offspring, there were insufficient degrees of freedom to resolve the dataset (Van Ooijen, 2009). IM was then selected over KW as it allows for analyses of a trait observed in multiple families to be combined over populations, increasing detection power (Van Ooijen, 2009).

Both regression and mixture model algorithms were tested in IM, with both algorithms identifying the same linkage groups and regions as significant. The regression approach was selected as this approach was less time intensive than the mixture model approach (Van Ooijen, 2009). A LOD test statistic, a fit for dominance for F2, a 5cM mapping step size, and a maximum of 20 neighboring markers were also selected as analysis parameters.

As singularity errors, or more than a single mathematical solution, were detected, CP populations were recoded to fit a pseudo-testcross approach, or a double haploid (DH) population type to try and rectify these errors (Van Ooijen, 2009). However, given the small family sizes in this study, this approach resulted in too large a loss of power for QTL detection and the dataset was reverted to a CP population encoding. To determine the appropriate significance threshold for QTLs permutation tests on a genome-wide level were conducted in MapQTL v. 6 for each trait (Van Ooijen, 2009).

3.2.8 Genome-wide Association Studies (GWAS)

Although QTL mapping approaches can be quite powerful in identifying associations between genotypes and phenotypes, they have greater power when families with a large number of offspring are available (Korte and Farlow, 2013). Alternatively, genome-wide association studies (GWAS) rely on unravelling linkage disequilibrium in markers in large, heterogeneous populations, and therefore draw their power from having a population of individuals from varying genetic backgrounds (Hayes, 2013, Korte and Farlow, 2013).

GWAS was undertaken to confirm detected QTL and to identify any additional genetic associations to growth and sex traits. A total of 388 phenotyped individuals and 6,163 markers were utilized in GWAS analyses for sex determination. Due to the identification of potential recording errors in generations 9 and 11, only the 216 samples from G10 were used for GWAS analyses for weight. To denote the accumulated effect of all SNPs, genetic relationship matrices were created in PLINK 1.9 beta (Chang et al., 2015, Purcell and Chang, 2017). These matrices were then used to perform a mixed linear model-based association analysis to detect markers under selection in GCTA v1.92.1beta5 (Yang et al., 2011, Yang et al., 2014). Two traits, sex and harvest weight, were examined with age as a quantitative covariate and earthen pond and generation as qualitative covariates. Sex was also used as a quantitative covariate when conducting GWAS for harvest weight. GWAS results were visualized using the R package qqman (Turner, 2014).

3.3 Results

3.3.1 Phenotypic Data

Of the 136 animals used in linkage mapping, phenotypic data was only available for 60 individuals. Of these, sex ratios were 51.7% female and 48.3% male. When all 388 animals used for GWAS were considered, sex ratios remained similar with 53.9% female and 46.1% male. Sex ratios for Family 10, used in QTL analysis, demonstrated the greatest disparity in sex ratios, with a ratio of 37.5% female to 62.5% male. Family 12, the second QTL family, exhibited a ratio of 50% female and 50% male.

Families 6 (314.4 ± 88.6g), 10 (290.5 ± 91.8), and 12 (311.5 ± 81.4g) demonstrated the largest average final weight, with Families 10 and 12 used in QTL analysis (Appendix 4). Families 9 (118.1 ± 43.3g) and the single individual in Family 13 with weight data available (74.2g) exhibited the lowest average final weight (Appendix 4). Weights in Generation 10 ranged from 57.8g to 512.5g, with 48 weight classes of 10g ranges identified (Appendix 5). Approximately 25% of individuals fell within the 190-230 weight classes, with the 200-210g weight class having the highest number of individuals (16) of all weight classes (Appendix 5). Over 32% of Generation 10 were categroized in a weight class of 300g or above, and approximately 6% were classified below 150g (Appendix 5).

3.3.2 Reference Mapping Families

All available families were used to create the female, male, and sex average linkage maps; however, only the two largest families with phenotypic data could be phased in JoinMap and used for QTL analysis (Families 10 & 12; Table 3.1). All phenotyped individuals were used for GWAS analyses.

The power for quantitative trait detection with the 216 individuals in G10 utilized of up to 1 (0.41 heritability, 0.05 Type 1 error rate, and 0.01 variance of the SNP-derived genetic relationships).

3.3.3 DNA Extraction, Genotyping, and SNP Filtering

Of the initial 486 samples, 99.4% passed the initial quality control with 121 individuals passing in G9, 216 individuals passing in G10, and 146 individuals passing in G11. A total of 6,163 high quality SNPs and 1,620 SilicoDArT markers were identified for subsequent linkage mapping, QTL analysis and GWAS.

3.3.4 Map Construction and Genome Coverage

While markers grouped to both available genome assemblies, only those grouped using Orenil1.1 had a sufficient number of linkages between markers to map (Fig. 3.1). Synteny comparisons between the two genome assemblies revealed many notable differences in marker groupings and positions (Fig. 3.2). Therefore, only the first genome assembly was utilized to draw comparisons between *de novo* mapping and mapping with known orders. A consensus map was generated as a conservative map containing markers that were mapped in both the *de* novo and genome-based mapping order methods. Within the consensus map, seven markers did not agree in linkage group placement between *de novo* and genome-based mapping order were classified as aberrant and removed from the linkage maps. Within the sex-specific maps, nine aberrant markers were removed from the female-specific map, and 31 were removed from the male-specific map.

The consensus map (sex average) grouped to 22 LGs, corresponding to the karyotype chromosome number of *O. niloticus.* It should be noted, that as LGs were named according to Orenil1.1 groupings, LGs range from LG 1 – LG 23, with LG 21 excluded from numbering as

this linkage group was collapsed into LG 16 after the first release of the assembly. Of these groupings, only 21 of the 22 LGs had enough meiotic events to map, with LG 10 not mapping (Appendix 6). The female-specific and male-specific maps also grouped to 22 LGs, with LG 10 not mapping in the female-specific map (Appendix 6) and LG 22 not mapping in the male-specific map (Appendix 6).

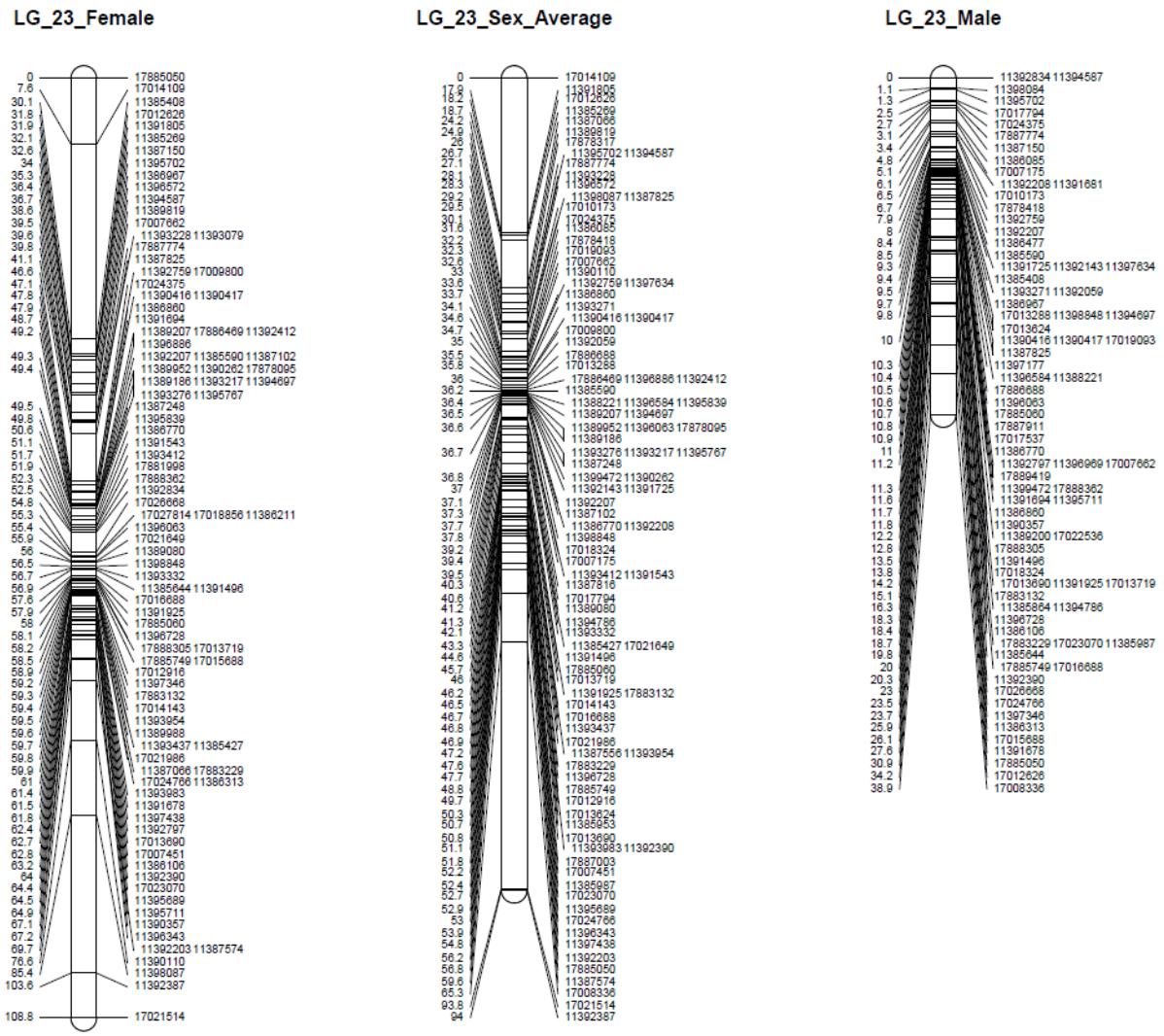


Fig. 3.6 LG 23 based on sex average, female, and male maps. Visualized maps for all linkage groups are provided in Appendix 6. Map file for all linkage groups (including marker, LG, and position order based on Kosambi's centimorgan is located in Digital Supplementary Material 1). LG 23 has been selected for display as it exhibited the strongest association with sex determination in both QTL and GWAS analyses.

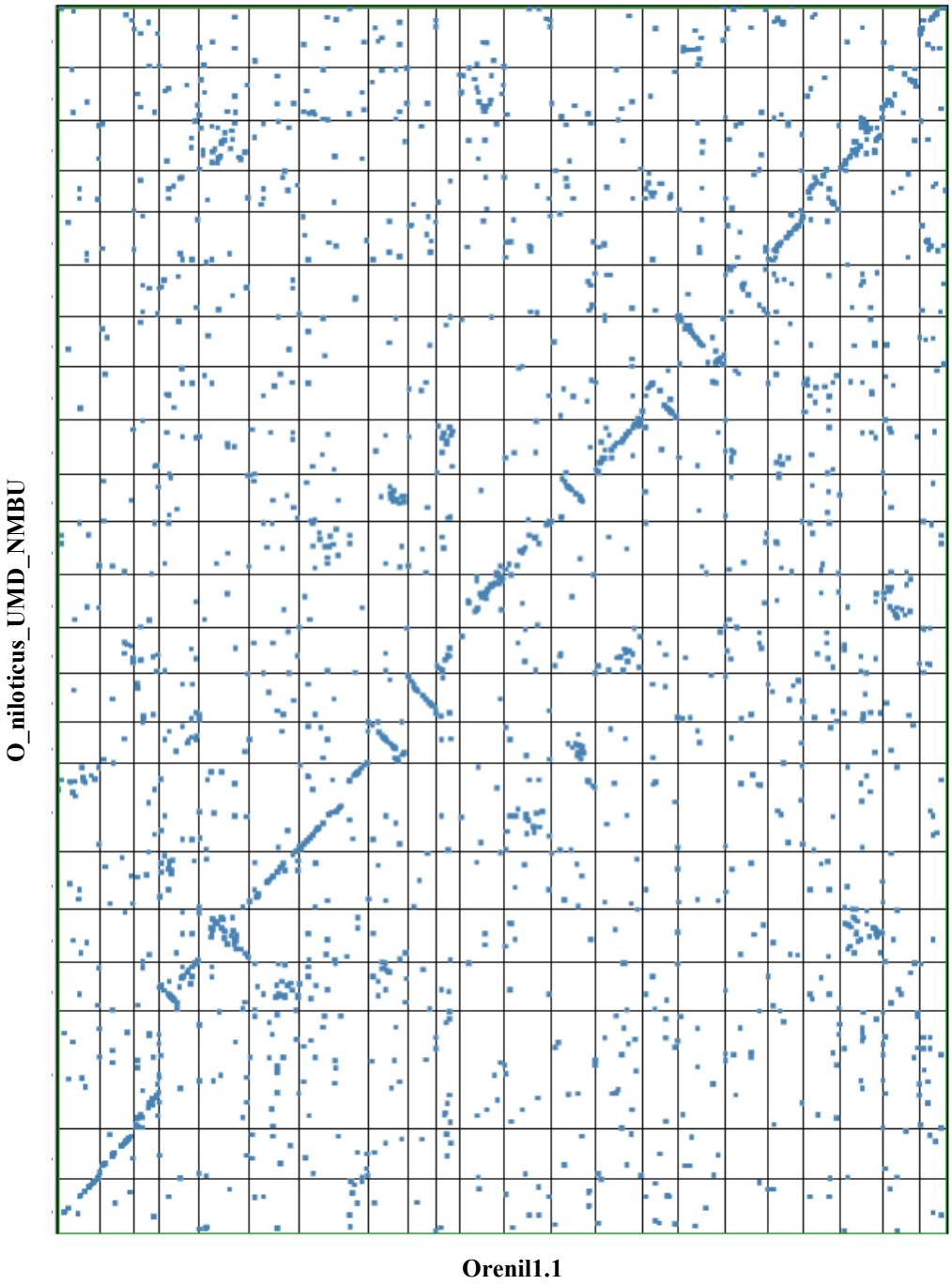


Fig. 3.7 Synteny graph comparing linkage group and marker order between the Orenil1.1 (GenBank assembly accession: GCA_000188235.2) and O_niloticus_UMD_NMBU (GenBank assembly accession: GCA_001858045.3) genome assemblies for *Oreochromis niloticus*. Each box represents a single chromosome, with box size dependent on the size of the mapped chromosome.

3.3.5 Segregation Distortion

Significant segregation distortions were detected in 348 of the 6,163 high quality SNPs available for mapping and fourteen of the thirty-two mapping parents following Bonferroni correction (mean family corrected alpha of 0.001). Of these, 115 SNPs mapped to the sex average map. These markers were from the largest seven families (8-17 offspring) with between 2-8 markers present in all 21 mapped linkage groups. As all maps were conducted using a population-based method which averaged and weighted LOD scores based on family and parental traces, these small family segregations were not considered to be strongly influencing calculations of mapping distances and were retained.

3.3.6 Map Metrics and Inter-chromosomal Analysis

A total of 2,399 markers were successfully mapped to their most likely positions within 21 of 22 linkage groups for the sex average map, whereas 2,197 markers mapped to the female-specific map and 2,125 markers to the male-specific map (Table 3.2).

Linkage groups within the sex average map span a total of 1,468.21 cM, with the female-specific map being 15.2% larger (1,688.90 cM), and the male-specific map being 3.3% smaller (1,419.23 cM; Table 3.2) in comparison to the sex average map. The number of markers per LG amongst all three maps ranged between 59-151 markers. The female-specific map produced the largest LG length on average (80.42 cM per LG), whereas the sex average map (69.91 cM per LG) and male-specific map (67.58 cM per LG) were similar in size. The female-specific map also had the largest average marker interval size (0.93cM) and the largest gap size (43.30 cM LG 22; Table 3.2).

Of the mapped markers, 36.8% were in common among all three maps (Appendices 6). The female-specific map had 71.1% of its markers in common with the sex average map, and 55.3% of its markers with the male-specific map (Table 3.3). Whereas, the male-specific map shared 66.4% of its markers with the sex average map, and 57.2% with the female-specific map. The remaining markers were unique to each map. A comparison of marker order and synteny of common markers amongst all three maps revealed that the average Spearman correlation for sex average map vs. female-specific map was 0.59, 0.68 for the sex average vs. male-specific map, and 0.41 for the female-specific vs. male specific map (Table 3.3, Figs. 3.3, 3.4, and 3.5) indicating that despite a high number of unique markers affecting placement, marker placement between groups was highly correlated.

Table 3.6 The number of markers, map size, average gap size, and largest gap size identified for each linkage map in the female, sex average, and male linkage maps.

LG	*Number of Markers*			*Map Size*			*Average Gap Size*			*Largest Gap Size*		
	Female	***Sex Average***	***Male***	***Female***	***Sex Average***	***Male***	***Female***	***Sex Average***	***Male***	***Female***	***Sex Average***	***Male***
1	89	115	84	94.23	90.00	76.80	1.18	0.83	0.96	25.55	14.09	10.14
2	68	94	115	49.65	48.82	66.90	0.79	0.52	0.87	6.47	4.94	12.49
3	69	63	59	100.67	80.01	80.48	1.68	1.40	1.49	22.99	18.30	12.94
4	137	151	129	78.10	54.16	82.09	0.71	0.45	0.86	22.66	4.17	13.95
5	100	113	94	102.89	80.71	77.87	1.13	0.75	0.86	16.24	21.25	14.66
6	105	121	92	76.14	68.19	61.52	0.77	0.60	0.74	12.31	8.53	13.69
7	147	119	140	58.97	89.12	67.42	0.49	0.80	0.56	7.75	25.35	9.69
8	113	120	117	50.85	48.24	52.75	0.54	0.41	0.57	3.57	3.32	4.96
9	89	75	83	84.54	41.54	60.62	1.13	0.60	0.84	16.51	5.60	8.16
10	--	--	64	--	--	103.35	--	--	2.58	--	--	21.83
11	139	145	104	91.72	90.75	87.29	0.91	0.76	0.87	23.14	21.38	10.35
12	108	109	106	80.98	63.78	50.43	0.89	0.61	0.50	12.55	7.72	3.04
13	88	128	106	90.60	65.82	97.14	1.18	0.55	1.03	13.96	4.48	12.71
14	111	126	119	78.75	51.33	46.51	0.80	0.42	0.43	11.62	7.58	5.96
15	132	113	95	69.69	53.59	59.58	0.73	0.50	0.65	11.94	2.99	5.85
16	118	123	113	94.41	69.41	57.63	0.93	0.59	0.65	23.05	10.47	7.49
17	115	130	92	73.83	59.85	67.49	0.72	0.49	0.77	12.47	6.85	13.96
18	75	104	93	98.75	81.53	51.02	1.41	0.84	0.65	39.56	22.80	10.66
19	77	110	110	39.29	55.18	70.96	0.58	0.54	0.76	4.46	4.89	7.62
20	115	138	129	89.72	101.70	62.44	0.88	0.76	0.55	20.03	21.86	9.08
22	104	100	--	76.28	80.50	--	0.84	0.82	--	43.30	17.08	--
23	98	102	81	108.85	93.98	38.93	1.24	0.99	0.56	22.42	28.51	4.76
All	2197	2399	2125	1688.90	1468.21	1419.23	--	--	--	--	--	--

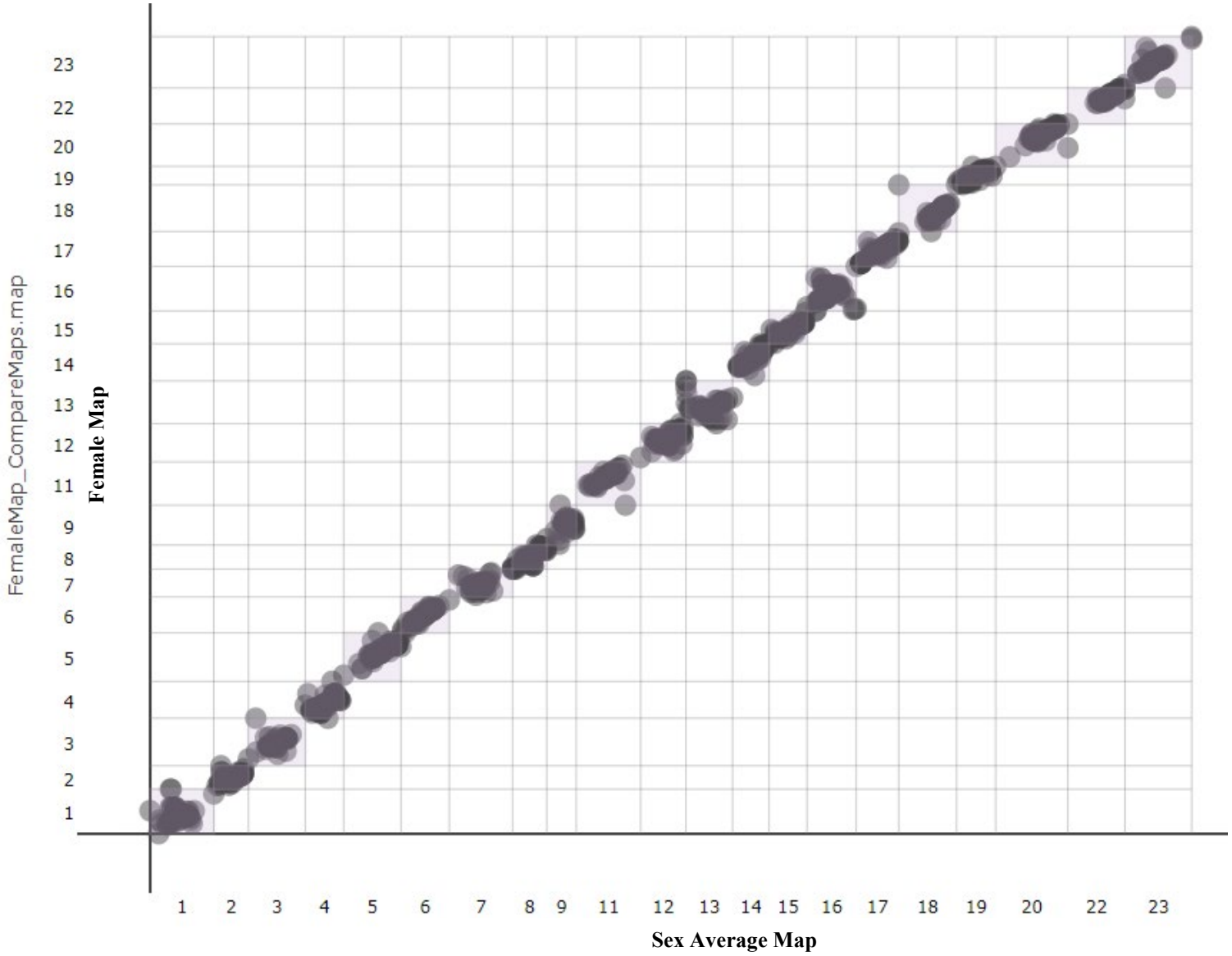


Fig. 8.3 Map order comparison of corresponding markers for the female vs sex average maps. Purple squares represent corresponding linkage groups between maps, markers are represented by grey dots with overlapping markers resulting in darker areas**.**

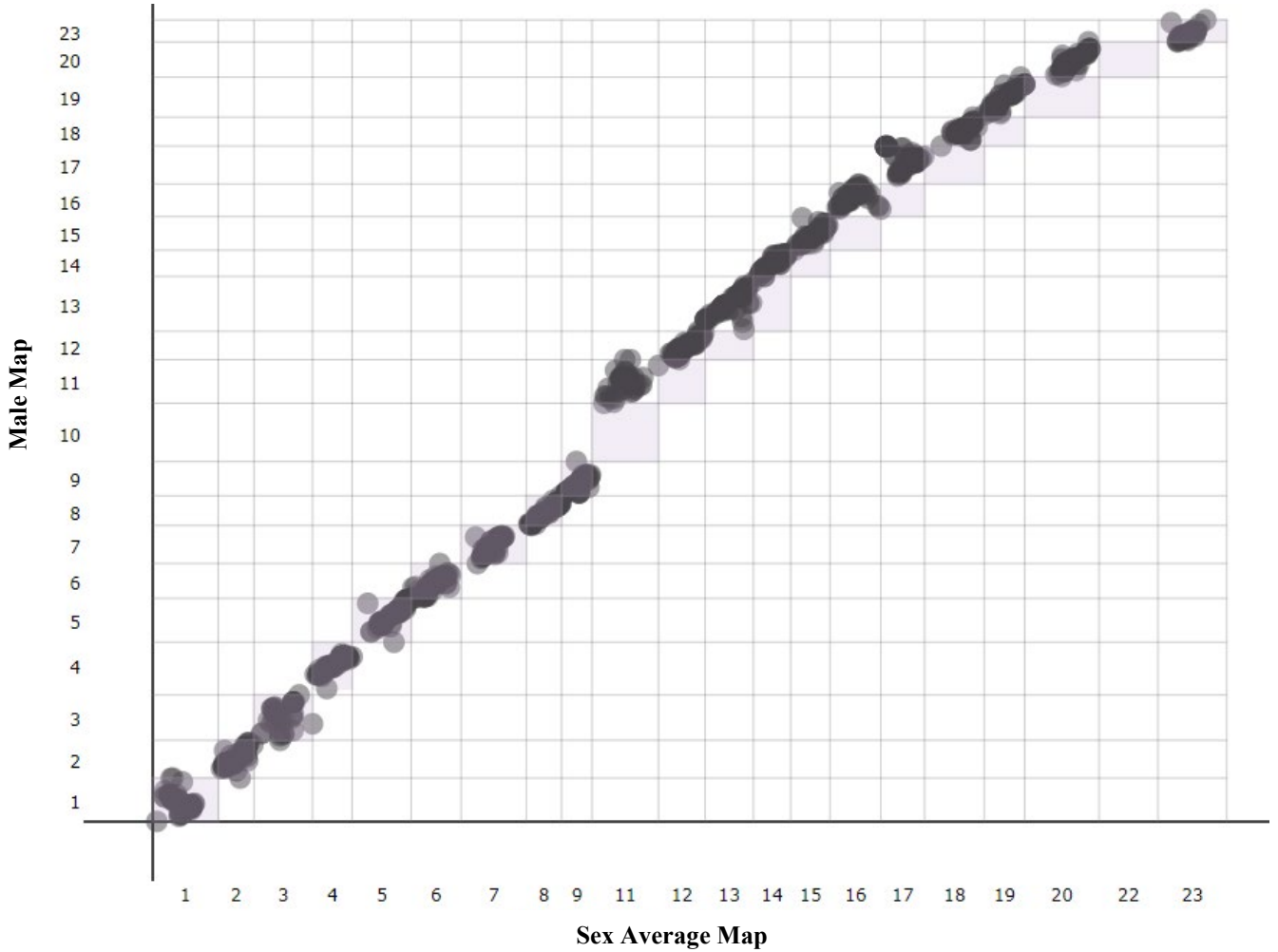


Fig. 3.9 Map order comparison of corresponding markers for the male vs. sex average map. Purple squares represent corresponding linkage groups between maps, markers are represented by grey dots with overlapping markers resulting in darker areas**.**

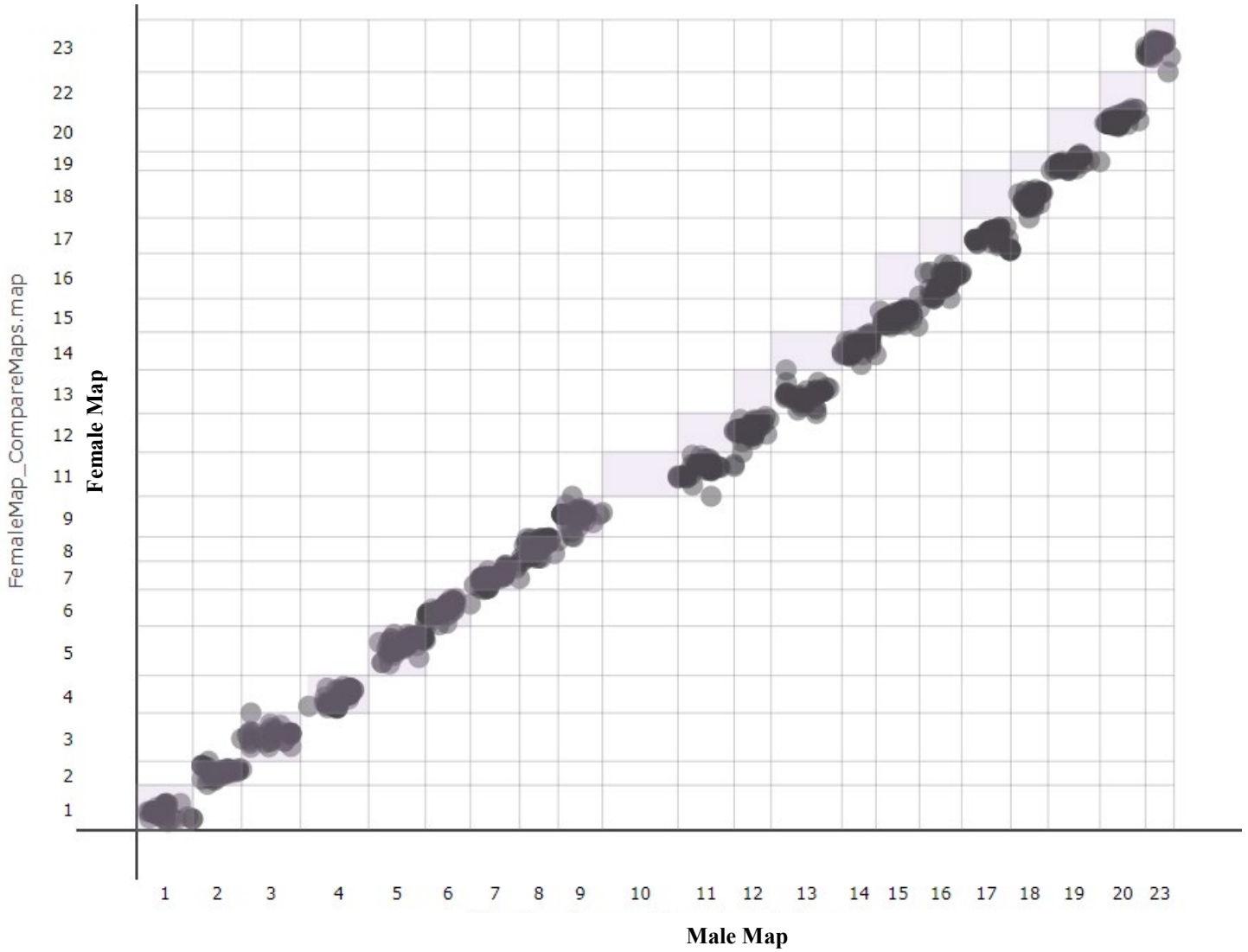


Fig. 3.10 Map order comparison of corresponding markers for the female vs. male map. Purple squares represent corresponding linkage groups between maps, markers are represented by grey dots with overlapping markers resulting in darker areas.

Table 3.7 Metrics of map comparison for the sex average, female, and male linkage maps.

	Number of Markers in Map			*Sex vs. Female Map*		*Sex vs. Male Map*		*Female vs. Male Map*	
LG	***Sex Average***	***Female***	***Male***	***Common Markers***	***Spearman Correlation***	***Common Markers***	***Spearman Correlation***	***Common Markers***	***Spearman Correlation***
1	115	89	84	72	0.09	66	0.53	53	0.08
2	94	68	115	51	0.57	72	0.79	46	0.16
3	63	69	59	43	0.44	35	0.28	34	0.1
4	151	137	129	96	0.6	94	0.74	89	0.68
5	113	100	94	80	0.8	73	0.88	64	0.65
6	121	105	92	71	0.93	62	0.82	56	0.79
7	119	147	140	77	0.28	78	0.75	83	0.56
8	120	113	117	83	0.7	76	0.9	72	0.58
9	75	89	83	54	0.05	55	0.77	61	0.01
10	-	-	-	-	-	-	-	-	-
11	145	139	104	112	0.77	68	0.16	65	0.1
12	109	108	106	76	0.46	71	0.89	70	0.28
13	128	88	106	66	0.05	78	0.75	53	0.18
14	126	111	119	76	0.88	84	0.87	69	0.57
15	113	132	95	81	0.86	66	0.82	66	0.6
16	123	118	113	97	0.26	92	0.54	83	0.54
17	130	115	92	85	0.76	62	0.22	48	-0.1
18	104	75	93	58	0.78	69	0.68	46	0.51
19	110	77	110	58	0.8	76	0.82	50	0.73
20	138	115	129	83	0.71	81	0.81	60	0.55
22	100	104	-9	68	0.88	-	-	-	-
23	102	98	81	75	0.77	53	0.66	48	0.54

3.3.7 Quantitative Trait Locus (QTL) Mapping

Suggested QTLs associated with sex were identified in LGs 8, 12, 14 and 23 for all three maps (sex average, female, and male) with the female map also identifying suggested QTLs in LGs 3 and 19 (Figs. 3.6, 3.7 and 3.8). Of these, only LGs 12 and 23 in the sex average map and LG 23 in the male map had QTL regions that were significant for both families were used in this analysis. As such, QTL regions of 46-65cM in LG 12 and 0-39cM in LG 23 of the sex average map and 27-40cM in LG 23 of the male map were considered to be putative QTLs (Figs. 3.6 and 3.8).

Suggested QTLs associated with growth were identified in LGs 2, 3, 8, 13, 14, and 18 in all three maps; LGs 4 and 5 in only the female and male map; LGs 6 and 19 in the sex average and female maps; LGs 17 and 20 in the sex average and male maps; LGs 7, 11 and 16 in the sex average map; and LG 12 in the female map (Figs. 3.9, 3.10, and 3.11). No QTL regions were identified as significant in both families and therefore not classified as putative QTLs.

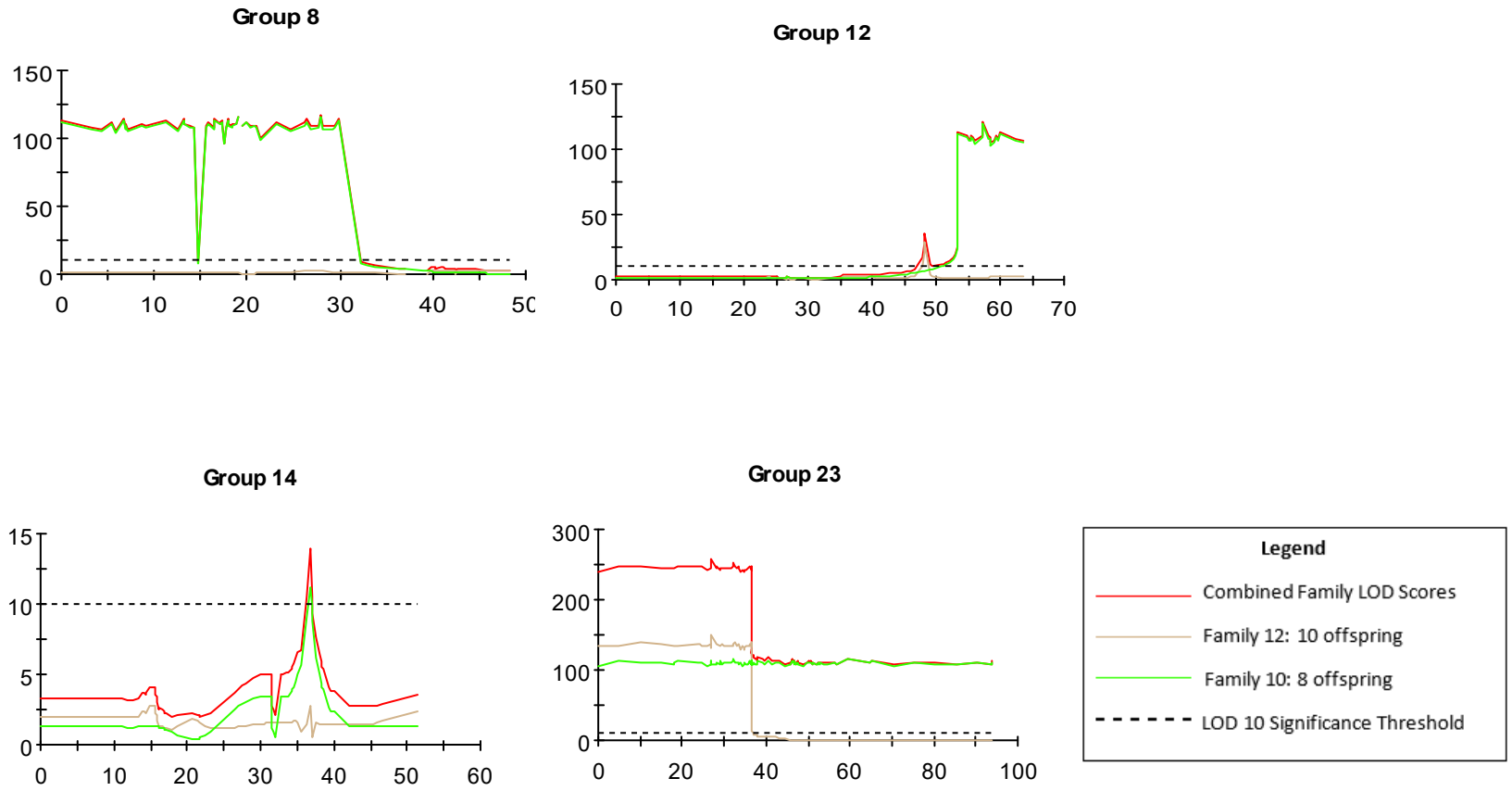


Fig. 3.11 QTL maps based on sex average linkage maps for markers associated with sex where the threshold for genome-wide significance is a LOD of 10.

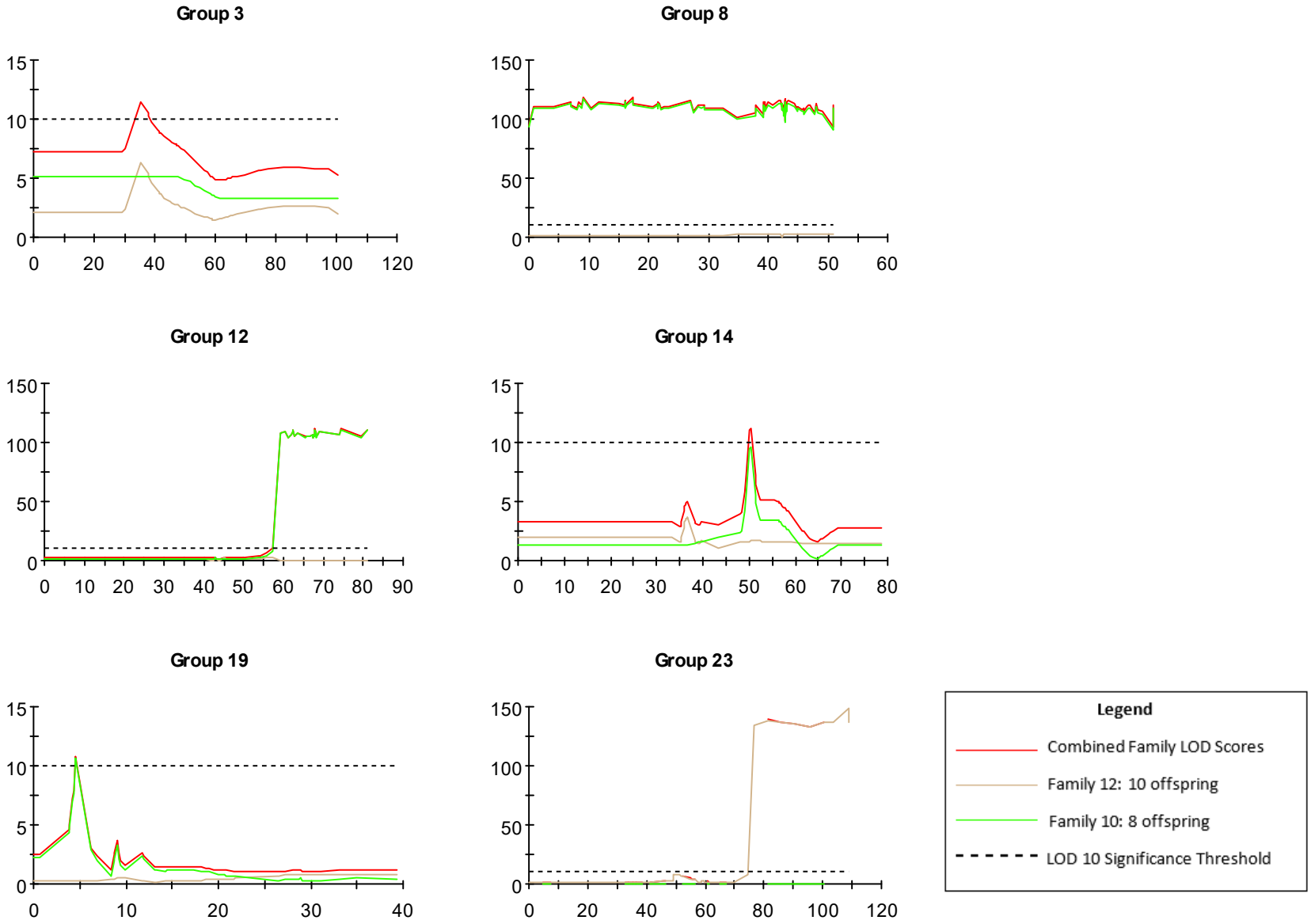


Fig. 3.12 QTL maps based on female linkage maps for markers associated with sex where the threshold for genome-wide significance is a LOD of 10.

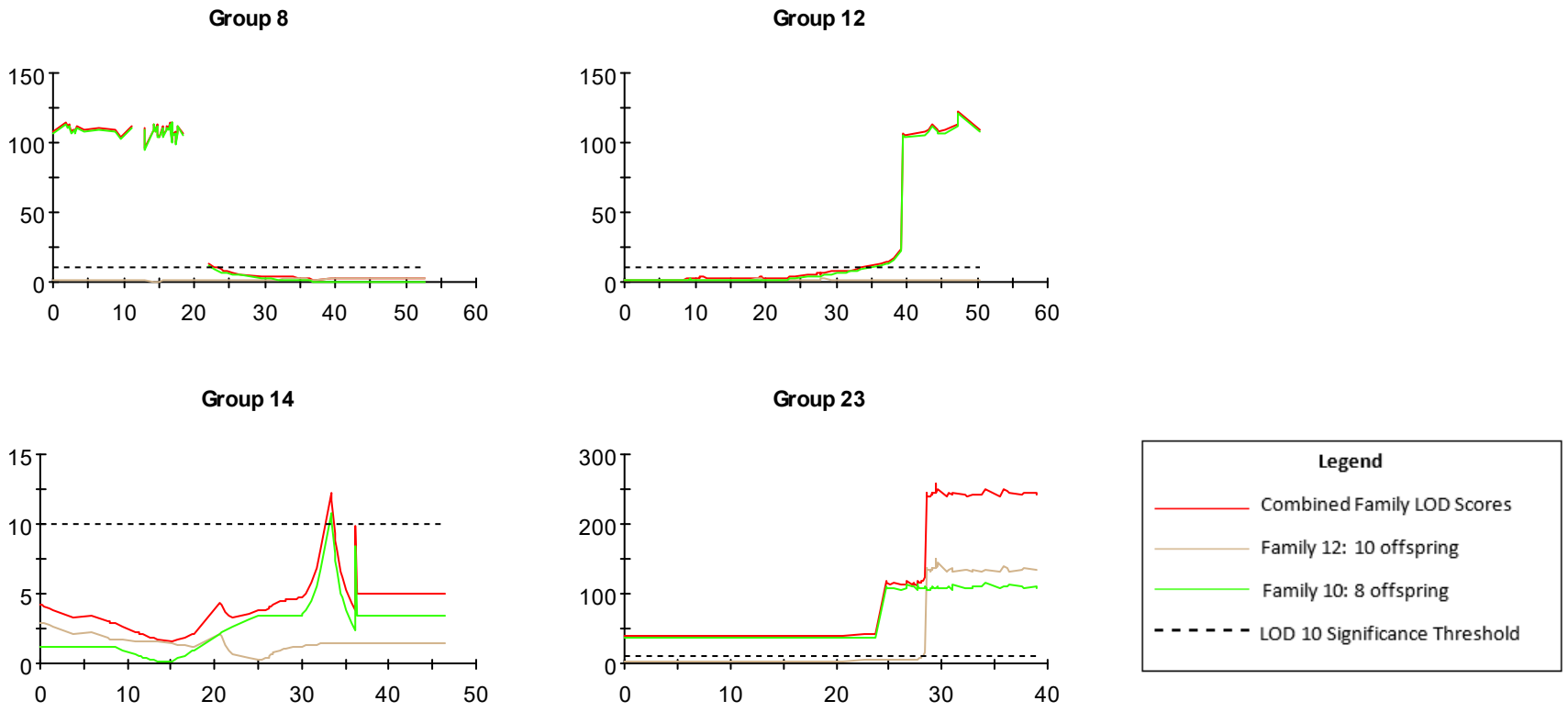


Fig. 3.13 QTL maps based on male linkage maps for markers associated with sex where the threshold for genome-wide significance is a LOD of 10.

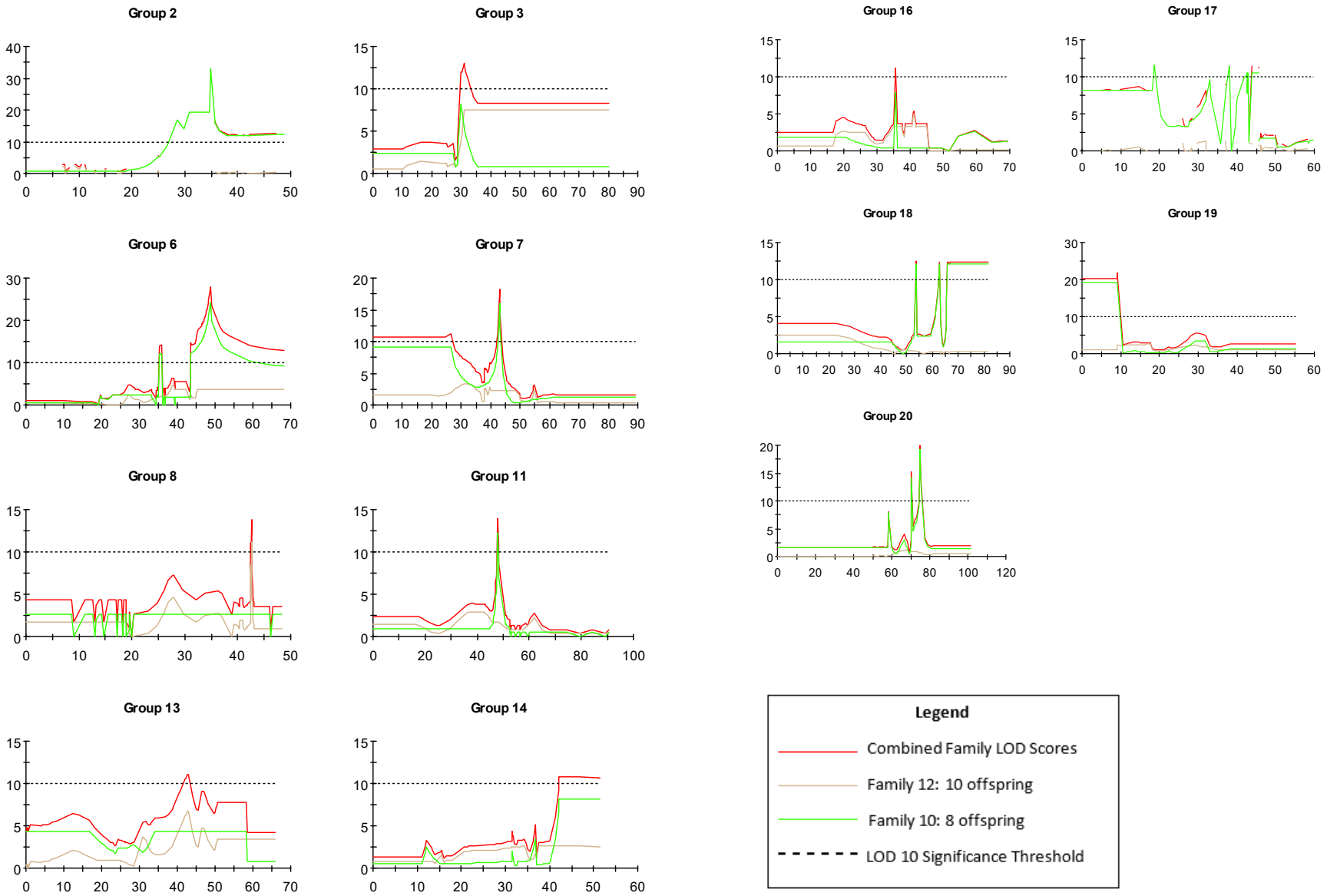


Fig. 3.14 QTL maps based on sex average linkage maps for markers associated with weight where the threshold for genome-wide significance is a LOD of 10.

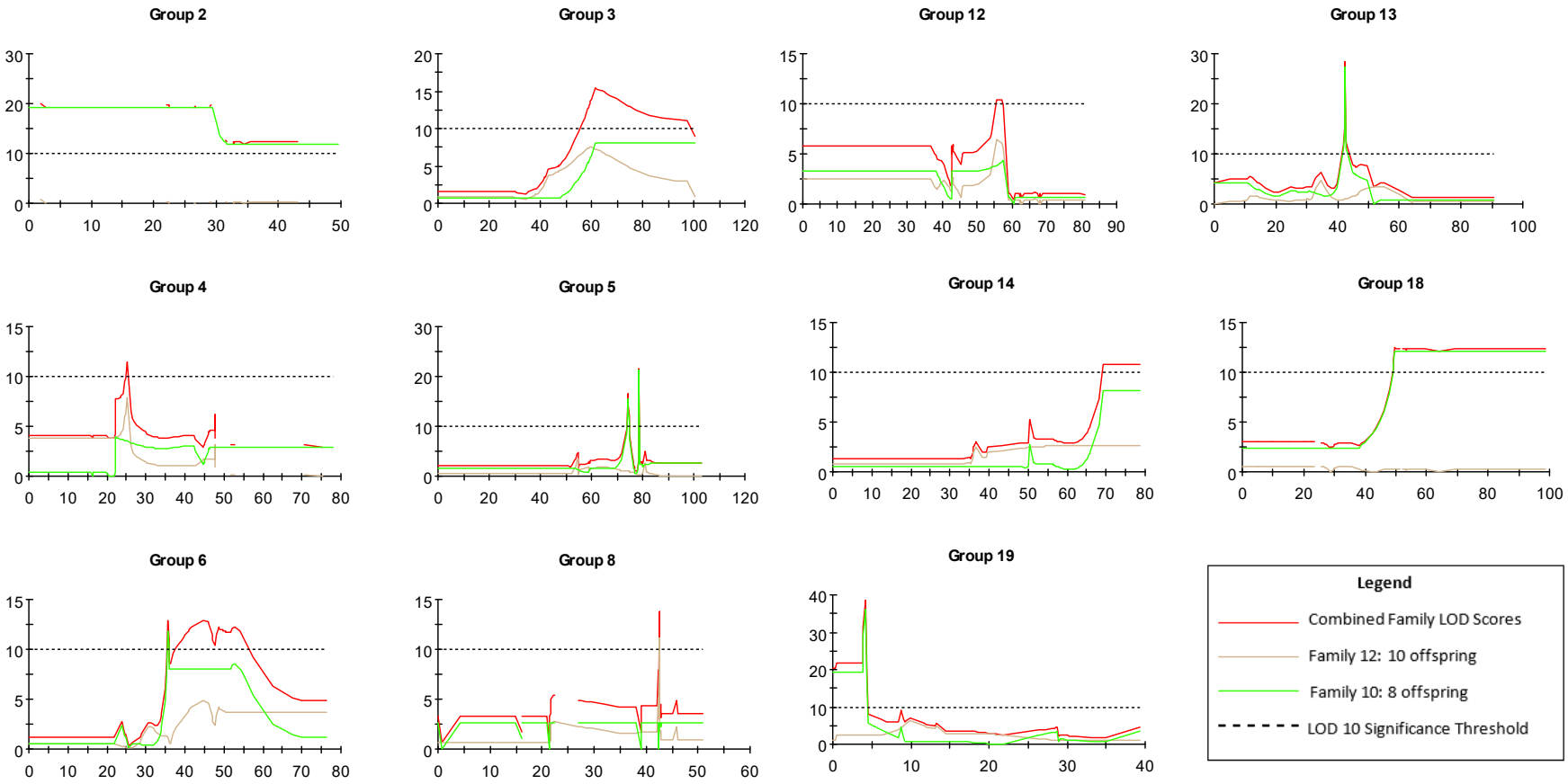


Fig. 3.15 QTL maps based on female linkage maps for markers associated with weight where the threshold for genome-wide significance is a LOD of 10.

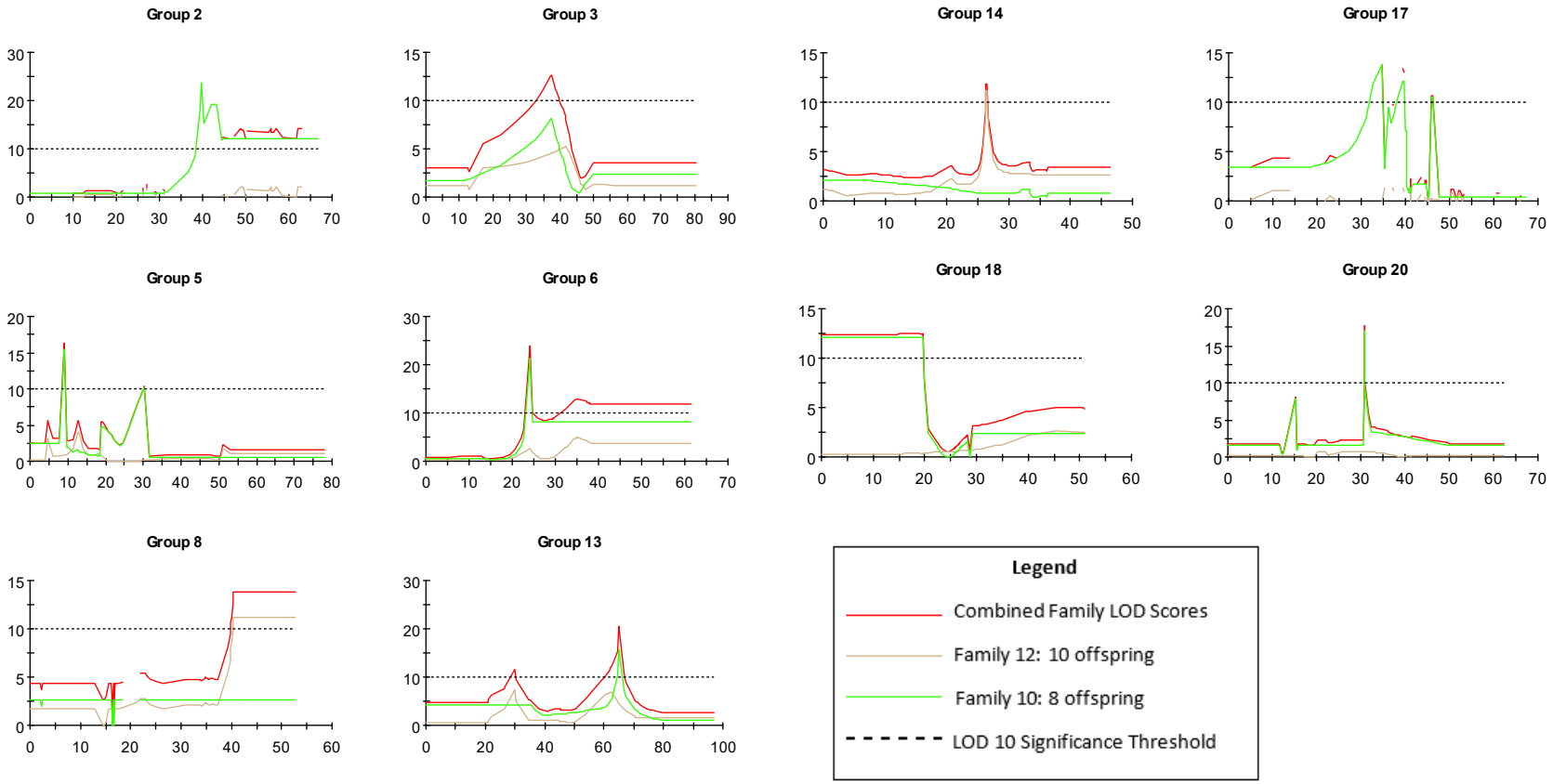


Fig. 3.16 QTL maps based on male linkage maps for markers associated with weight where the threshold for genome-wide significance is a LOD of 10.

3.3.8 Genome-wide Association Studies (GWAS)

Nine markers were identified as being associated with sex in the AS ($p \leq 1x10^{-5}$), with seven of these markers (17886688, 17007662, 11397634, 11396584, 11393271, 11391681, and 11388221) mapping to LG 23 in at least one of the available maps (sex average, female-specific, and male-specific; Figs. 3.12, 3.13, and 3.14). The sex average map placed six markers (three at $p \leq 1x10^{-5}$ and three at $p \leq 1x10^{-10}$) on LG 23 (Fig. 3.12). In the female-specific map, a single marker associated with sex was located on LG 23 ($p \leq 1x10^{-10}$; Fig. 3.13). In the male-specific map, six markers were associated with sex on LG 23 (three at $p \leq 1x10^{-5}$ and three at $p \leq 1x10^{-10}$; Fig. 3.14). Segregation by sex was also investigated within these markers, and while there were no absolute differences in segregation, all nine GWAS markers associated with sex showed one sex with higher proportions of A or B alleles than the other (Fig. 3.15). In particular, markers 11396584 (males 69.8% AB, females 96.7% AA), 11393271 (males), and 11388221 males 49.7% AB, females 77.0% AA) had the most definitive segregation between males and females as well as the highest significance for GWAS values in all maps (Figs. 3.12, 3.13, 3.14, and 3.15). No markers were found to be significantly associated with growth rate (Appendices 7, 8, and 9).

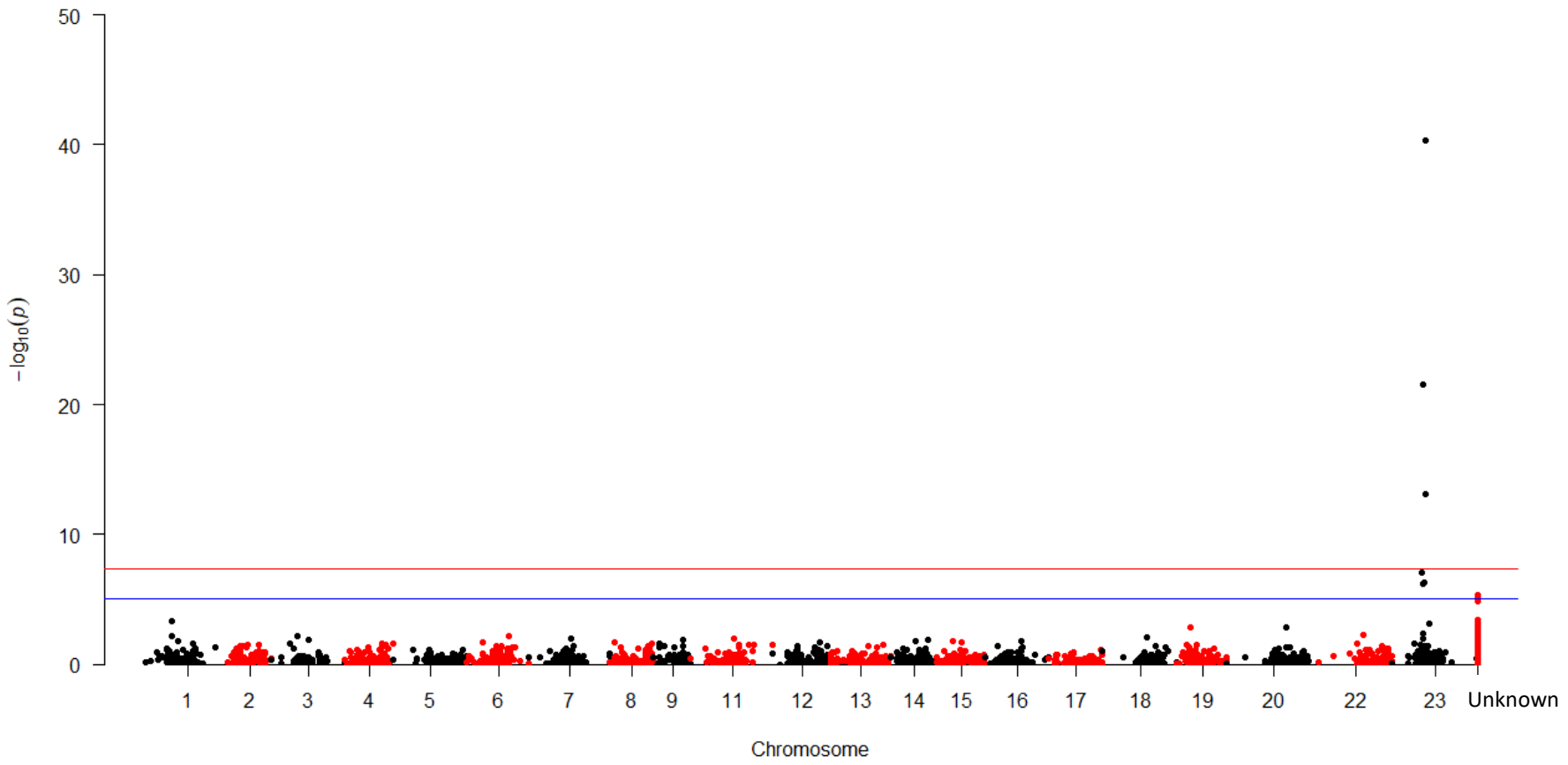


Fig. 3.17 Manhattan plot based on sex average linkage map for markers associated with sex. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line.

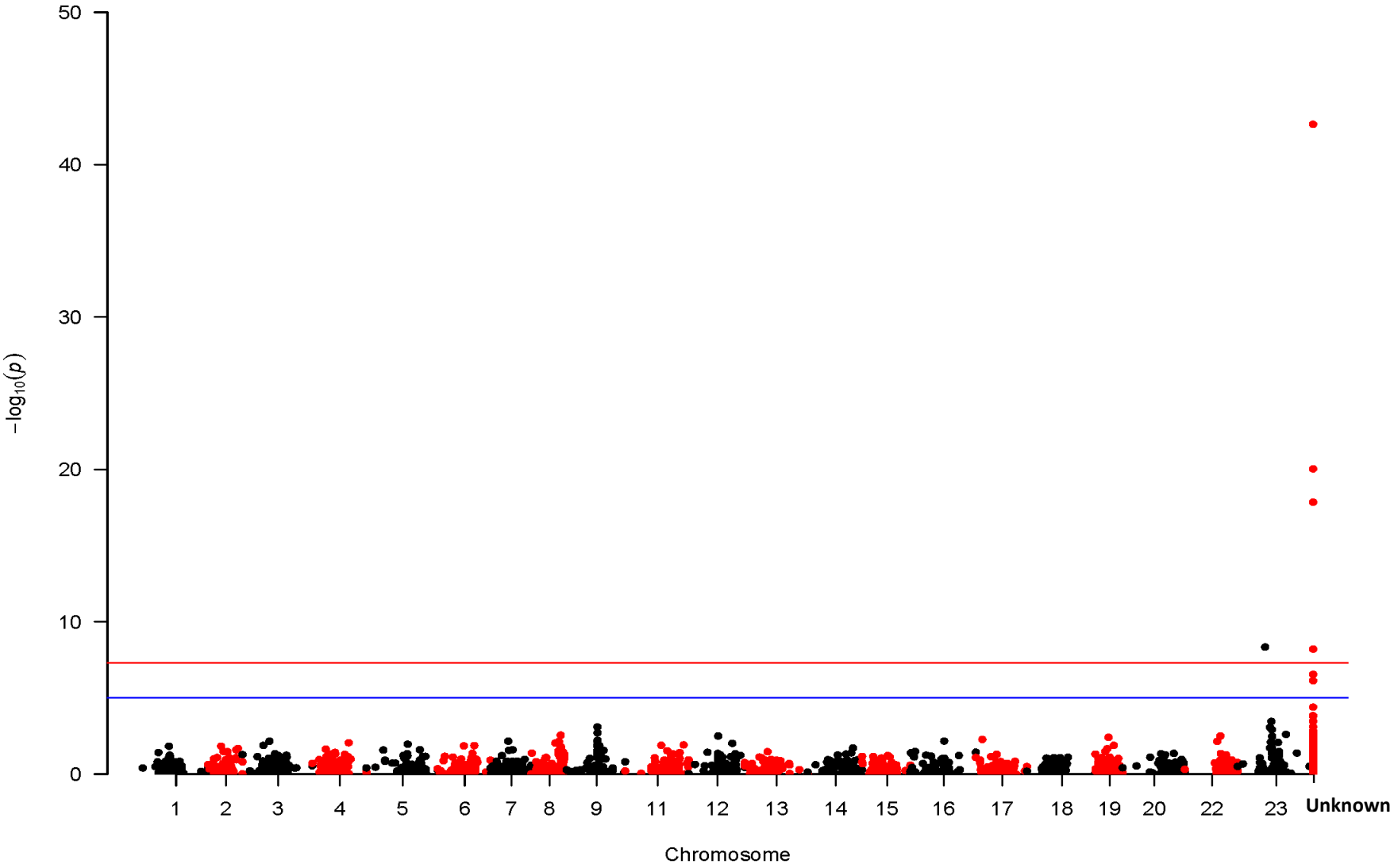


Fig. 3.18 Manhattan plot based on female linkage map for markers associated with sex. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line

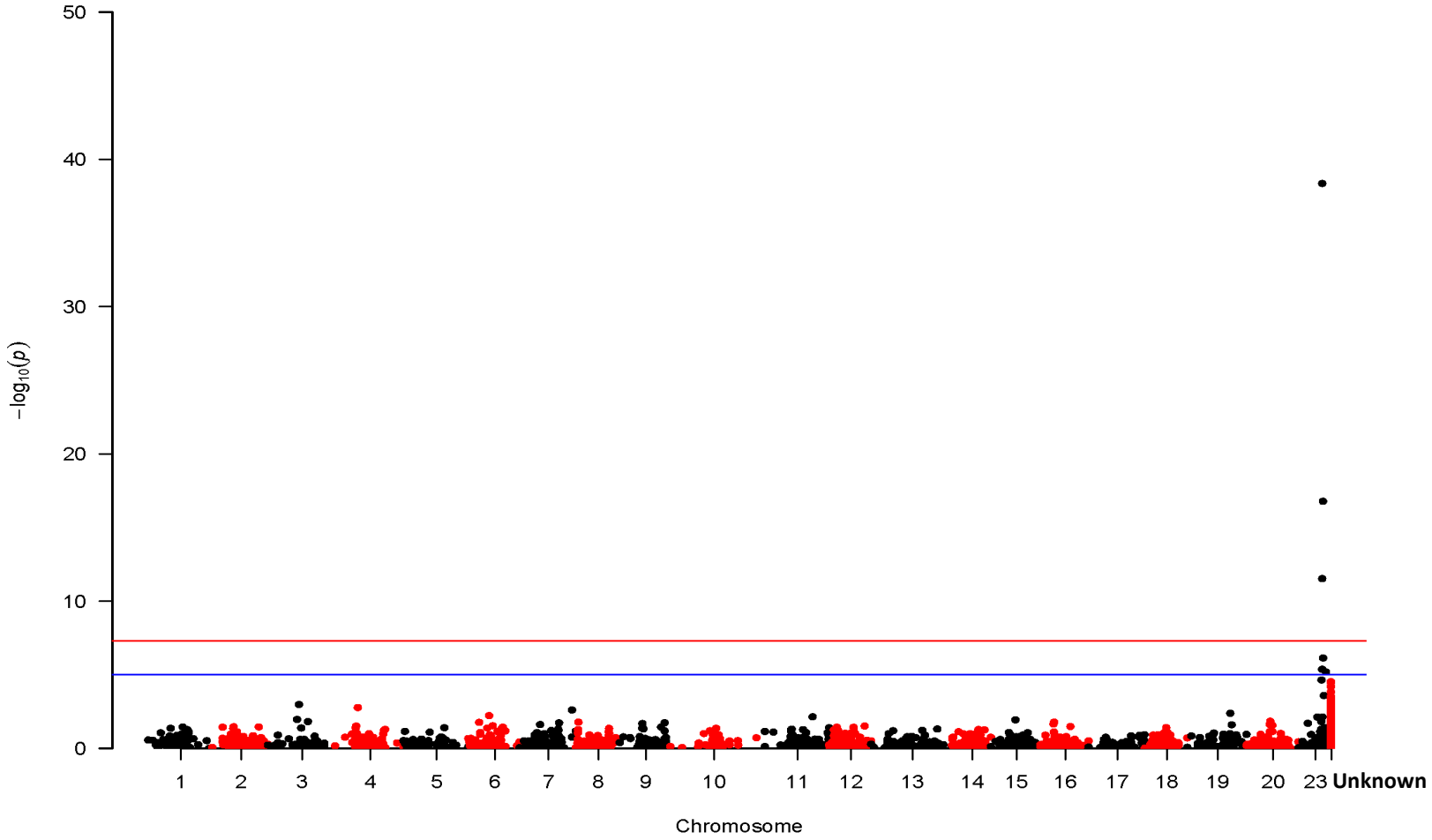


Fig. 3.19 Manhattan plot based on male linkage map for markers associated with sex. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line.

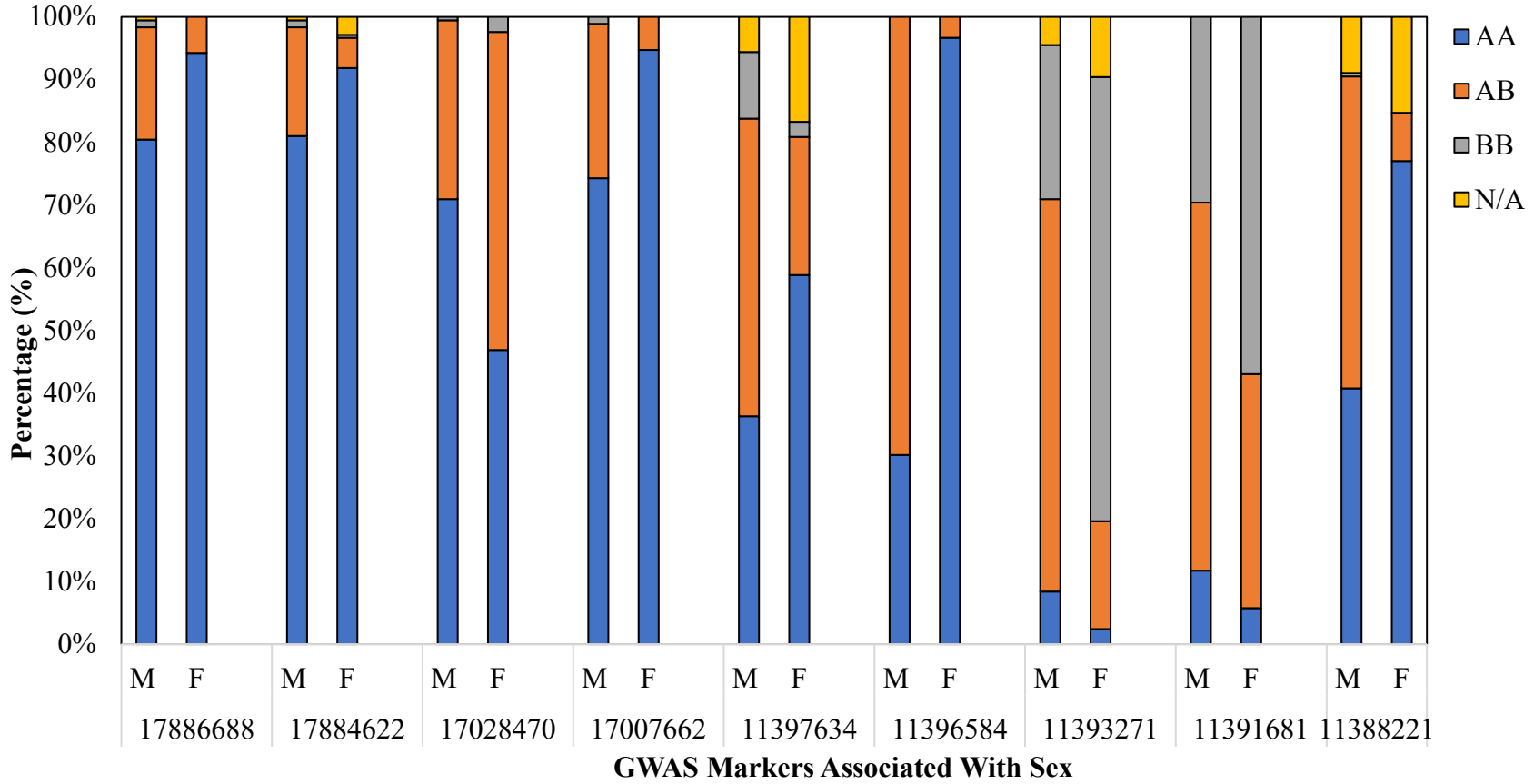


Fig. 3.20 Segregation of important GWAS markers associated with sex. A comparison of the proportion of phenotypic male and female progeny exhibiting respective genotypes; AA (blue), AB (Orange), BB (grey), and missing data (yellow).

3.4 Discussion

This study produced the first line specific linkage map for the commercially important Abbassa strain of Nile tilapia to investigate quantitative trait loci associated with sex and weight. This linkage map is one of the first population-based genetic linkage maps using small families (≤17 offspring) and phase unknown data. Due to the atypical construction of this map, independent and unique maps were created based on female lines, male lines, and the sex average. A total of 2,399 markers were successfully mapped to a sex average map, 2,197 to a female map, and 2,125 to a male map. All maps and map orders were validated by the reference genome for Orenil 1.1. Phenotypic data for sex and final weight at harvest were then utilized to determine regions of the genome associated with sex and weight. This study provides evidence of QTLs for sex and weight on LGs 3, 8, 12, 14, 9, and 23, and 2, 3, 4, 5, 6, 7, 8, 11, 12, 13, 14, 16, 17, 18, 19, and 20 respectively. GWAS validated QTL analysis for LG 23's association with sex determination.

Three linkage maps, a sex average, female line, and male line, were created for the AS. These maps all placed between 2,125-2,399 SNPs, with all three linkage maps representing some of the denser maps available for Nile tilapia (Guyon et al., 2012, Joshi et al., 2018, Kocher et al., 1998, Lee et al., 2005, Palaiokostas et al., 2013). The sex average map length observed in this study is within 2 cM of that observed in the largest available linkage map (40,186 markers; Joshi et al., 2018) and over 292 cM longer than the map created by Palaiokostas et al. (2013) with 3,802 markers mapped. Both available reference genomes were tested in the construction of the AS linkage map, but it was found that the AS only mapped with Orenil 1.1 (NCBI, 2017) ordering and not O_niloticus_UMD1 (Joshi et al., 2018). A comparison of these two genome assemblies to one another revealed major discrepancies. These differences may be attributed to the source of the animals used for the assembly. It is important to note that unpublished research by the

WorldFish Center and affiliated researchers has indicated that the AS is not pure *O. niloticus*, but has hybridized with blue tilapia, *O. aureus* (Grobler, 2017). This, in addition to the assembly discrepancies, illustrates the importance of strain specific linkage maps and genome assembles.

The present study found that population-based linkage mapping using small, two generational families is a viable method for map construction. This method required the creation of unique sex average, male and female maps which resulted in differences in the size of each map and the markers that were mapped. Despite the high number of unique markers that had the potential to affect marker placement, there was high correlation between marker placements across all three maps (sex average, female, and male). Additionally, all maps had high concordance with the marker groupings and orders within the Orenil 1.1 (NCBI, 2017) genome assembly for Nile tilapia. Such accuracy indicates strong potential for this method to be utilized in other species which may be limited by family size, either through life histories, sampling constraints, or unforeseen challenges (including, but not limited to, funding and sampling limitations due to pedigree errors).

Although small family sizes also limited QTL detection power, QTLs were still identified within the AS. A comparison of putative sex QTLs (in this case, those regions which were identified as significant in both families in QTL analysis) and GWAS results were found to be in agreement with one another with regions in LG 23 associated with sex determination in *O. niloticus* (sex average map: 0-93.8 cM; female map: 36.5-108.8 cM; and male map: 0-38.9 cM) supported by both QTL and GWAS analysis across all maps and believed to be associated with a major QTL segregating with sex. These results agree with previous studies which found LG 23 (annotated to the Orenil1.1 genome assembly or the O_niloticus_UMD_NMBU genome assembly, with both assemblies anchored using the same NCBI RefSeq automated eukaryotic genome annotation

pipeline) to be strongly linked to sex determination (Cáceres et al, 2019, Joshi et al., 2018, Palaiokostas et al., 2015). Segregation by sex for all nine markers identified in GWAS studies did not find any fixed alleles, although allele frequencies did vary substantially between males and females. Three markers in particular, 11396584, 11393271 and 11388221, had the clearest distinction in segregation and had the highest significance in GWAS values.

Though sex determination is considered to be a conserved trait in many organisms, this is not the case in teleost fishes, like tilapia, where both genetics and environment can trigger the mechanisms underlying sex determination (Mank and Avise, 2009, Nivelle et al., 2019). The leading idea for *O. niloticus* is that it embodies sex determination system of a heterogametic male (XY) and homogametic female (XX) (Mair et al., 1991). However, there is evidence that *O. niloticus* may have a polygenic sex determining system, in which sex determining loci are located on autosomes in addition to sex chromosomes (Baroiller et al., 2009, Baroiller et al., 1995, Wessels et al., 2014). This system is made more complex as morphological differences are absent between X and Y chromosomes in Nile tilapia (Campos-Ramos et al., 2001, Carrasco et al., 1999, Lee et al., 2004) and environmental factors, like temperature, have been shown to override the genetic sex determining system (Baroiller et al., 2009, Baroiller et al., 1995, Wessels et al., 2017).

To date, no study-including the present study-has clearly assigned a linkage group to the sex chromosomes for Nile tilapia (Cáceres et al., 2019, Conte et al., 2017, Eshel et al., 2012, Eshel et al., 2011, Lee et al., 2003, Palaiokostas et al., 2015). The strongest regions associated with sex in the AS were detected in LG 23, agreeing with a study conducted on the GIFT strain, a line also established using Egyptian stock annotated to the Orenil1.1 genome assembly (Grobler, 2017). Suggestive and putative QTLs associated with sex were also identified in five other linkage

groups, indicating that *O. niloticus* may exhibit a more complex sex-determining system (Khanam, 2017). This may be in part because the XY chromosomes are still in the early stages of differentiation with no strong morphological differences in any chromosome pair that would identify the X and Y chromosomes (Campos-Ramos et al., 2001, Carrasco et al., 1999, Lee et al., 2004). In organisms with X and Y heterogametic sex determination, parts, if recombination halts in the Y chromosome, this can eventually result in the evolution of a smaller Y chromosome due to genetic degeneration (Bergero and Charlesworth, 2009). However, direct comparisons between sexes were made difficult as unique maps were created for male and female lines in this study. It should be noted that recombination rates varied greatly between male and female maps for LG 23, LG 10, and LG 22. In particular, the male map had lower recombination rates and a shorter length for LG 23 compared to the female map, LG 22 only had enough informative meiotic events to map in the female map, and LG 10 only had enough informative meiotic events to map in the male map. It is suggested that these three LGs, particularly LG 23 as it has shown evidence of being associated with sex, be examined further as candidates for the XY chromosome.

The putative sex QTL region identified in LG 12 of the sex average map, and the suggested sex QTLs in LG 14 (all three maps) and LG 19 (female map) are novel for Nile tilapia. Based on genome annotation, previous studies have only identified LG 1 (Conte et al., 2017, Eshel et al., 2011, Palaiokostas et al., 2015), LG 2 (Eshel et al., 2011), LG 3 (Eshel et al., 2011, Palaiokostas et al., 2015), LG 6 (Eshel et al., 2011), LG 8 (Lee et al., 2003) and LG 20 (Palaiokostas et al., 2015), in addition to LG 23, as containing QTLs associated with sex. Of these linkage groups, LGs 3 (female map) and 8 (sex average, female, and male maps) were found to have suggested QTLs, with only one family exhibiting the association, in this study.

The intricate sex determining system of *O. niloticus* is further complicated by Nile tilapia's readiness to hybridize with other tilapia species (D'Amato et al., 2007, Deines et al., 2014, Lovshin, 1982, Meier et al., 2019). This includes *O. aureus* which is thought to have a ZW sex determining system (Campos-Ramos et al., 2001). This is of high relevance for the AS which is estimated to be comprised of 90% *O. niloticus* genetic material and 10% *O. aureus* (Grobler, 2017). A suggestive QTL was identified on LG 14, which in another study has been shown to affect the interspecific interaction in reproduction between these two species (Shirak et al., 2019). Additionally, LG 3 has been associated to sex determination in both *O. niloticus* and *O. aureus* (Eshel et al., 2011, Lee et al., 2004, Palaiokostas et al., 2015) and was hypothesized as a potential LG of interest in relations to sex determination in the AS by a previous study due to the AS's hybridization with *O. aureus* (Grobler, 2017). Considering that sex determination may be polygenic in tilapia species and the substantial variation in results in previous studies, it is feasible that the novel putative QTL identified in LG 12 along with the suggestive QTL in LG 19 are associated with sex determination either in *O. aureus,* or in the reproductive interaction between *O. niloticus* and *O. aureus*. However, to unravel its significance and whether it is truly associated with *O. aureus*, additional and more targeted experimentations are required.

No putative markers or genome regions were associated with weight at harvest; however, a total of 16 linkage groups were identified with suggested QTLs. Previous studies on Nile tilapia and tilapia species have associated LG 1, LG 3, LG 7, LG 10 (Liu et al., 2014), LG 12 (Lin et al., 2016), LG 13, LG 19 (Liu et al., 2014) , LG 20, and LG 22 (Lin et al., 2016) to weight, suggesting that weight is a polygenic trait and likely varies between families, populations, and species (Lin et al., 2016). This is supported by data from other tilapia species which have indicated that growth rate is a polygenic trait (Cnaani et al., 2004). Furthermore, similar to this

study where only six of the ten linkage groups identified with weight QTLs observed across our three maps, the majority of weight QTLs in previous studies have been shown to be sex-specific (Liu et al., 2014). The novel linkage groups associated with weight QTLs (LG 2, LG 3, LG 8, LG 13, LG 14, and LG 18) may be attributed to either the *O. aureus* genome or the interaction between the combination of the *O. niloticus* and *O. aureus* genomes. However, to date, there have been relatively few studies on weight QTLs in purebred Nile tilapia, blue tilapia, and Nile tilapia x blue tilapia hybrids, and weight should be examined more closely in these species to better understand both the effects of hybridization and selective breeding on this trait of commercial interest on the genome.

Typically, *O. niloticus* grows to a larger size than *O. aureus;* however, first generation hybrids of *O. niloticus* x *O. aureus* outperform both of their purebred parents in regards to final harvest weight (El-Hawarry, 2012). As 10% of the AS genome is estimated to be derived from *O. aureus* and the remaining 90% from the intended *O. niloticus* (Grobler, 2017), it is unknown what the long-term genetic consequences of the initial hybridization event(s) are. With other hybrid examples that initially exhibited hybrid vigor, it was lost in subsequent generations due to a loss of heterogeneity (Johansen-Morris and Latta, 2006). It is feasible, that the incorporation of the *O. aureus* genome into the AS may also explain the modest genetic gain (3.8-7.0%; Rezk et al., 2009) recorded within the AS compared to other purebred Nile tilapia selection lines (7.1-15.0%; Eknath and Acosta, 1998, Ponzoni et al., 2011). To determine if this hybridization has influenced genetic gain in the AS, more targeted studies are required.

3.5 Conclusions

Breeding programs have become an integral methodology to improve production in aquaculture species. In cases such as Nile Tilapia, with numerous selective breeding programs and the ability

to hybridize with other tilapia species, it is important to produce strain specific linkage maps to improve accuracy in subsequent QTL and GWAS analyses. In addition to demonstrating the importance of strain specific maps, the feasibility of using population-based linkage mapping constructed with regression models as an alternative to family-based linkage mapping was shown, particularly when a reference genome is available. Such techniques have implications for non-model species, particularly where fewer offspring are available, or in cases where sampling is not ideal. This study identified putative QTLs segregating with sex; including, LG 23 which was validated using both QTL and GWAS and agreed with previous studies, and a novel association with LG 12 based on QTLs. The suggested QTL in LG 14 may be associated with reproduction and sex determination as a result of hybridization between *O. niloticus* (XY) and *O. aureus* (ZW). This illustrates the complexity of sex determining systems in tilapia species and the need for more studies to be conducted to understand the effects of hybridization on the genome. Given that weight is a polygenic trait in Nile tilapia and that no putative QTL or GWAS regions were associated with weight within this study, marker assisted selection for weight is not currently feasible for the AS; however, genomic selection may be viable. Selective breeding in aquaculture is a field that is moving rapidly. Results presented here are integral to ongoing studies investigating potential genetic selective breeding avenues for Nile tilapia which may include the identification of additional associations that better explain the phenotypic variance of trait/s, and help industry move towards genomic selection.

CHAPTER 4: COMPARING GENOMIC SIGNATURES OF SELECTION BETWEEN THE ABBASSA STRAIN AND EIGHT WILD POPUALTIONS OF NILE TILAPIA (*Oreochromis niloticus*) IN EGYPT

4.1 Introduction

As aquaculture production continues to increase, so do the number of species undergoing domestication (currently estimated at 598 species; FAO, 2018), where domestication is defined here as the adaptation of an organism from the wild to a captive environment (Price, 1984). These adaptations can be a combination of genetic changes that occur over generations through selective breeding for desirable traits like size (Argue et al., 2002, Eknath and Acosta, 1998), disease resistance (Moss et al., 2012, Robinson and Hayes, 2008) and color (Hossain et al., 2011, Wan et al., 2017), but also include adjustments to a captive environment, such as reduced antipredator behaviors and aggression (Johnsson et al., 1996, Robinson and Hayes, 2008).

The four main genetic processes that affect animals during domestication are founder effects, selection, genetic drift, and inbreeding (Andueza-Noh et al., 2015, Clutton-Brock, 1992, Ladizinsky, 1985, Mignon-Grasteau et al., 2005, Ollivier, 2002); however, the extent of their effects on the genome can vary. In general, the consequences of inbreeding and genetic drift are widespread and can be observed throughout the genome, whereas selection tends to act differentially across the genome depending on the genetic architecture of the trait (Burke et al., 2005). These micro-evolutionary processes need to be taken into consideration when trying to identify how an organism's genome is being affected by domestication.

One way to understand the genetic consequences of domestication and to identify signatures of selection is to compare population genetic metrics between captive and wild populations (López et al., 2019, Simmons et al., 2006). Recent advances in high-throughput whole genome sequencing has enabled the cost-effective development of genome-wide markers for many non-model species. Such technological developments have enabled researchers to not only harness increased power in identifying the extent to which genetic processes like selection, genetic drift, and inbreeding affect a genome, but also identify specific regions of the genome that have responded to such processes (Carter et al., 2008, López et al., 2019, Scandura et al., 2011). Evaluating the genetic differences between wild and domestic populations can therefore also help determine genomic regions associated with domestication and desirable market traits, identify wild populations that exhibit these traits, identify local adaptations in wild populations, and detect escapees and their potential impact on local populations.

In 2002, the Abbassa strain (AS) of Nile tilapia (*Oreochromis niloticus*) was initiated by the WorldFish Center in an effort to increase aquaculture production of this species in Egypt (Ibrahim et al., 2013, Rezk et al., 2009). Its purpose was to provide a genetically diverse population based on the local strain of Nile tilapia that could be selectively improved for growth. Subsequently, the AS was created from four Egyptian populations (three wild: Zawia, Abbassa, and Aswan; one hatchery: Maryout). The production of AS is currently restricted to the Nile Delta; however, WorldFish and the Egyptian government plan to disseminate the AS line throughout Egypt.

To date, genetic diversity studies have found that wild Nile tilapia populations have evidence for sub-structuring in Egypt, particularly between populations in the Nile delta in Upper Egypt compared with populations in the Lower Egyptian portion of the Nile River (Hassanien and

Gilbey, 2005, Hassanien et al., 2004). However, due to the age of the studies, possible translocations and the availability of improved genetic technologies, updated investigations into the genetic structure of these populations using high density, genome-wide markers are required to determine the current status of wild population structuring.

This study investigated the genetic diversity, population genetic structure, and evidence for signatures of selection related to domestication in the AS compared to wild Egyptian Nile populations. This information can then be used to understand the impact disseminating the AS may have on wild stocks, as well as understand if targeted breeding in the AS has resulted in signatures that may be indicative of domestication.

4.2 Material and Methods

4.2.1 Sampling and DNA Extraction

4.2.1.1 Wild Population Sampling

Fin clips from 400 Nile tilapia were collected from eight wild populations (Aswan, n = 50; Manzala Lagoon, n = 50; Kanata, n = 50; Lake Idku, n = 50;Damietta, n = 50; Lake Brulus, n = 50; Rosetta, n = 50; and Asyut, n = 50) along the Nile River, Egypt. Of these, Aswan was one of the four sites from which individuals were sampled to create the domesticated Abbassa Strain in 2002. The samples for an individual location were obtained over a distance of approximately 1 to 175 km. Fin clips were taken from fish obtained directly from commercial fishing boats. Samples were preserved in 70% ethanol and submitted to Diversity Arrays Technology (DArT) in Canberra, Australia, for DNA extraction and high throughput genotyping by sequencing using proprietary DArTseq^{TM} technology (https://www.diversityarrays.com). To obtain purified DNA,

extractions were conducted using commercially available extraction kits (Promega, Qiagen; Lind et al., 2017).

4.2.1.2 Abbassa Strain Population Sampling

Fin clips from 486 samples were collected from generations 9-11 of the AS [121 individuals from generation 9 (G9); 216 individuals from generation 10 (G10); and 146 individuals from generation 11 (G11)]. DNA extractions and genotyping were conducted by Diversity Arrays Technology (DArT) as described in Lind et al. (2017).

4.2.2 Library Preparation and Sequencing

DNA extractions and genotyping were conducted by Diversity Arrays Technology as described in Lind et al. (2017) and in Chapter 2. To ensure complete digestion and a uniform range of fragment sizes, all samples were checked using an agarose gel. Any samples which displayed downshifted bands after digestion during DArTseq library preparation were removed. These downshifted samples exhibited a lower amplicon range than expected when compared to other samples and are not ideal for a consistent genotype assay. A total of eight downshifted samples were not included within the sequencing effort. Additionally, a minimum of 15% random technical were included in all genotyping batches for quality control.

4.2.3 Quality Control and Initial SNP Calling

DArT's proprietary marker calling algorithm DArTsoft14 was used to call SNPs (Lind et al., 2017), implemented in the KDCompute framework (http://www.kddart.org/kdcompute.html). Samples from wild locations were then co-analyzed by DArT alongside 483 samples from three generations of the AS, which had already been processed using DArTseq^{TM} technology as part of a previous experiment (Nayfa et al., 2020).

A total of 19,505 SNP markers were identified across all 875 samples and were filtered using a custom Python script adapted from DartQC (https://github.com/esteinig/dartqc) and CD-HIT-EST (Li and Godzik, 2006). Briefly, samples with greater than 50% missing data were removed from the dataset and individual genotypes calls made with fewer than five reads were silenced. Genotypes with a count comparison, or the comparison of read counts between REF and SNP alleles, were silenced if they fell between 0.05 and 0.1, where < 0.05 is considered to be homozygous and > 0.1 is considered to be heterozygous (https://github.com/esteinig/dartqc). SNPs were then filtered if they had an average replication statistic of less than 90%, a call rate less than 50%, and a minor allele frequency (MAF) of less than 1% in at least one population. The clone ID sequences from which SNPs were called and clustered together at 95% similarity using CD-HIT-EST (Li and Godzik, 2006). Within each cluster, the SNP with the highest minor allele frequency (MAF) was retained to ensure a more even representation of the genome. A total of 9,827 high quality SNPs and 821 samples (90.9% of collected samples) were retained for all downstream analyses.

4.2.4 Population Structure

4.2.4.1 Broad Scale Population Structure

To determine broad-scale population differentiation across the eight wild locations and three generations of the AS, two separate clustering models (the allele frequencies correlated model and the allele frequencies independent model) were utilized within a Bayesian cluster population structure analysis in STRUCTURE 2.3.4 (Falush et al., 2003, Falush et al., 2007, Hubisz et al., 2009, Pritchard et al., 2000). In order to avoid inappropriate clustering due to *K* being set too small, *K* was set from 1 to 12, so that the maximum clustering possible was larger than the number of putative populations (Kalinowski, 2011). Three repeat runs were performed for each

K (1-12), with a burn-in period of 5,000 iterations followed by 50,000 final iterations using the admixture model and no prior probabilities for cluster membership. Both clustering models yielded near identical results. The optimal number of population clusters, *K,* was determined using an ad hoc statistic *Delta K* (Δ*K*). Δ*K* is the degree of change in the log probability of data between successive *K* values, which was calculated using Structure Harvester (Earl and vonHoldt, 2012, Evanno et al., 2005). To ensure that any structuring observed in the wild populations was not biased by the inclusion of individuals from a domesticated line, analyses with the same parameters were repeated on only the eight native sampling locations testing a *K* of 1 to 9.

4.2.4.2 Fine Scale Population Structure

Fine-scale population genetic structuring across all eight wild sampling locations and the three AS generations was assessed using pairwise relationships based on identity-by-state (IBS) distance calculated in Plink v.1.9 (Purcell and Chang, 2017, Purcell et al., 2007). These relationships were then visualized using mutual k-nearest neighbor graphs in the NETVIEW pipeline v.1.1 at kNN values between 1 and 100 (Neuditschko et al., 2012, Steinig et al., 2016).

To test if any identified genetic structuring followed an isolation-by-distance model of population divergence, Mantel's test for correlation between genetic distance (F_{st}) and physical distance (m) was conducted in the R package *adegenet* using 10,000 permutations in the *mantel.randtest* function (Jombart, 2008, Jombart and Ahmed, 2011). Genetic distance was calculated using a Euclidean method based on Angular distance in the *adegenet* function *dist.genpop* (Jombart, 2008, Jombart and Ahmed, 2011). Geographic distances were calculated based on the shortest distance between two point according to the 'Vicenty (ellipsoid)' method calculated using the R package *geosphere* (Hijmans et al., 2017).

4.2.5 Signatures of Selection

4.2.5.1 Population Outlier Analysis

To identify outliers (including loci which are being influenced by selective processes), two independent methods were utilized: Arlequin 3.5.2.2 (Excoffier and Lischer, 2010) and BayeScan 2.1 (Foll, 2012, Foll and Gaggiotti, 2008). For comparisons between the two groups (wild Nile tilapia and the domesticated AS Nile tilapia), only those candidate outliers that were jointly identified between programs were categorized as putative outliers. Outlier analyses within Arlequin 3.5.2.2 were based on a hierarchical island model with 20,000 simulations, 50 simulated groups, and 100 demes simulated per group (Excoffier and Lischer, 2010). AMOVA computations were conducted using a pairwise difference method with no Gamma correction (Excoffier and Lischer, 2010).

Outlier analyses within BayeScan 2.1 were based on a neutral model with 1:10 prior odds, 20 pilot runs consisting of 5,000 iterations each, followed by 100,000 iterations with a burn-in-length of 50,000 iterations as recommended by Foll (2012). To establish whether a neutral or selection model was in effect for each SNP the ratio of posterior probabilities, Bayes factors (BF) were calculated. A Jeffrey's interpretation of "strong" BF (p-*value* ≤ 0.05) to "decisive" BF (p-*value*-≤ 0.01) was then utilized to identify outliers and ascertain which model the posterior odds favored (Foll, 2012). For markers which fell under a selection model, positive alpha values were then used to identify markers that were under diversifying or directional selection, whereas negative alpha values were used to identify those markers under background, or balancing, selection (Foll, 2012). For pairwise comparisons of populations, only BayeScan was used since 1) the hierarchical method utilized in Arlequin required the use of multiple populations per

grouping which was not present in the current data, and 2) the majority of outliers in genetic clusters identified by BayeScan were also identified by Arlequin 3.5.2.2 (approximately 70%).

To test for the normality of markers, quantile-quantile plots (QQ-plots) with a 95% confidence interval were constructed in the R package GWASTools v. 3.1 (Gogarten et al., 2012) for the full marker set, as well as the neutral marker sets (Gondro et al., 2013, Hayes, 2013). To validate the outlier selection criteria selected (i.e. markers jointly identified by both BayeScan and Arlequin), QQ-plots using the two different neutral marker sets (one with all identified outliers removed and one with only jointly identified outliers removed) were created. How well the data fitted the assumption of normality was then compared between both datasets, and only those jointly identified were retained. Comparison of these datasets allowed the validity of identified outliers to be established.

4.2.5.2 Genomic Regions Under Selection

Raw clone sequences from which SNPs were identified during the DArTseq process were annotated to the available genome assembly for *Oreochromis niloticus* (GenBank Assembly Accession: GCA_00188235.2; Orenil1.1) using a custom Perl script based on NCBI CGI BLAST interface with a 70% minimum sequence identity (Heller-Uszynska et al., 2011). The Orenil1.1 genome assembly was used instead of the more recent O_niloticus_UMD_NMBU assembly as it was in greater agreement with the linkage maps created for the Abbassa Strain (Chapter 3). In order to detect genomic regions under selection and determine the significance of outliers Manhattan plots were created using qqman v 0.1.4 in R v 3.5.2 (R Core Team, 2018, Turner, 2014).

4.2.6 Population Diversity Statistics

To determine the amount of genetic diversity within each sampled population (wild and AS), observed (H_o) and expected (H_e) heterozygosity in addition to the number of polymorphic markers within a population were calculated in ARLEQUIN 3.5 (Excoffier and Lischer, 2010). Heterozygosity and the number of polymorphic markers were examined across scenarios with different amounts of missing data (all markers; 5% missing allowed per SNP within individual populations; and 50%, 25%, and 5% missing allowed per SNP across the entire dataset). Additionally, average multi-locus heterozygosity (MLH) for each population was computed using the R package inbreedR (Stoffel et al., 2016). Private SNPs per population were calculated using the R package PopGenKit v.1.0 (Paquette and Paquette, 2011). To determine the level of differentiation amongst populations, pairwise and global F_{st} values were calculated in ARLEQUIN 3.5 (Excoffier and Lischer, 2010). Levels of inbreeding per sampling location and time point were examined using the inbreeding coefficient (F_{is}) calculated in ARLEQUIN 3.5 using 1,000 permutations (Excoffier and Lischer, 2010). Hardy-Weinberg equilibrium (HWE) was calculated in ARLEQIUN 3.5 using 1,000,000 Markov chain steps and 100,000 dememorization steps (Excoffier and Lischer, 2010, Waples, 2014). Effective population size in each native location was calculated using the linkage disequilibrium method (LDN_e) in NeEstimator V2.01 (Do et al., 2014).

4.3 Results

4.3.1 Population Structure Analysis

4.3.1.1 Broad Scale Population Structuring

The ad hoc ΔK statistic indicated evidence for two major genetic clusters within the dataset (Appendix 10). This distinction was supported by STRUCTURE admixture analysis whereby the domesticated AS generations formed one genetic cluster and the eight wild populations comprised the second cluster (Fig. 4.1 and Appendix 10). With a K of two, the admixture model used in STRUCTURE assumes that each individual has ancestry from only one or both of these genetically distinct clusters (Lawson et al., 2018). Given this, every individual from the AS shares genetic material with the wild Nile tilapia. This is reflected in the minimal population structuring identified between AS and wild sampling locations identified by pairwise F_{st} values (F_{st} = -0.008-0.058; Appendix 11). The largest genetic distance was observed between the two most southern wild sampling locations (Asyut and Aswan) and the AS (F_{st} = 0.045-0.058; Appendix 11).

Within the wild sampling locations, one individual from Rosetta was found which was more closely related to the AS than to the wild genetic cluster (Fig. 4.1). There are two individuals from Damietta, two individuals from Kanater, and one individual from Aswan which also had a higher proportion of shared ancestry with the AS than expected based on the other individuals in the wild genetic cluster (Fig. 4.1).

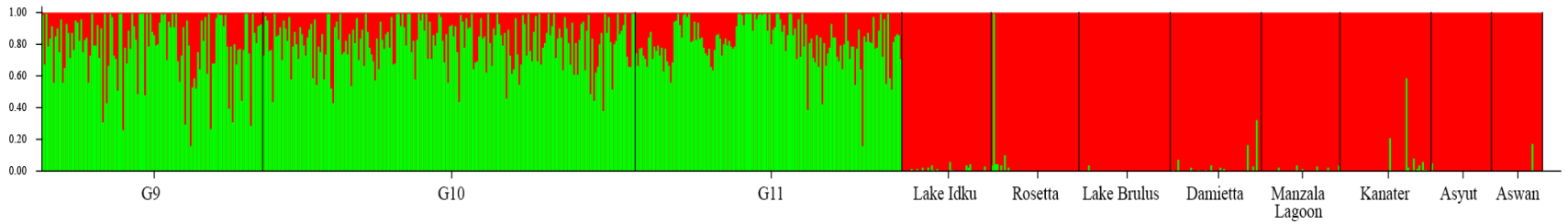


Fig. 4.21 Broad-Scale Population Structure. Structure plot of the three AS generations and the eight wild sampling locations at ΔK = 2. The wild sampling locations are ordered via geographical distance order.

When the eight wild locations were examined separately, the ΔK statistic identified a total of four weakly separated genetic clusters (Fig. 4.2; Appendix 12). While each sampled location showed evidence of all four genetic clusters within them, the proportion of these genetic clusters changed along the northern to southern gradient of the Nile River. The two most southern populations (Asyut and Aswan) exhibited the greatest difference in admixture ratios compared to Lake Idku, Rosetta, Lake Brulus, Damietta and Manzala Lagoon (Fig. 4.2). Kanter displayed the largest shift between the northern and southern sampling locations (Fig. 4.2). This was supported by pairwise F_{st} values which revealed no subpopulation structuring amongst the wild populations. The greatest F_{st} was between the northernmost population (Lake Brulus) and the southernmost population (Aswan; $F_{st} = 0.021$; Appendix 11).

Individuals showing an independent genetic cluster (green) in Fig. 4.2 were the same as individuals which displayed a greater association with the AS in Fig.4.1. This pattern suggests these individuals are possibly escapees (Rosetta) or subsequent offspring (Kanater, Damietta, and Aswan) of the AS.

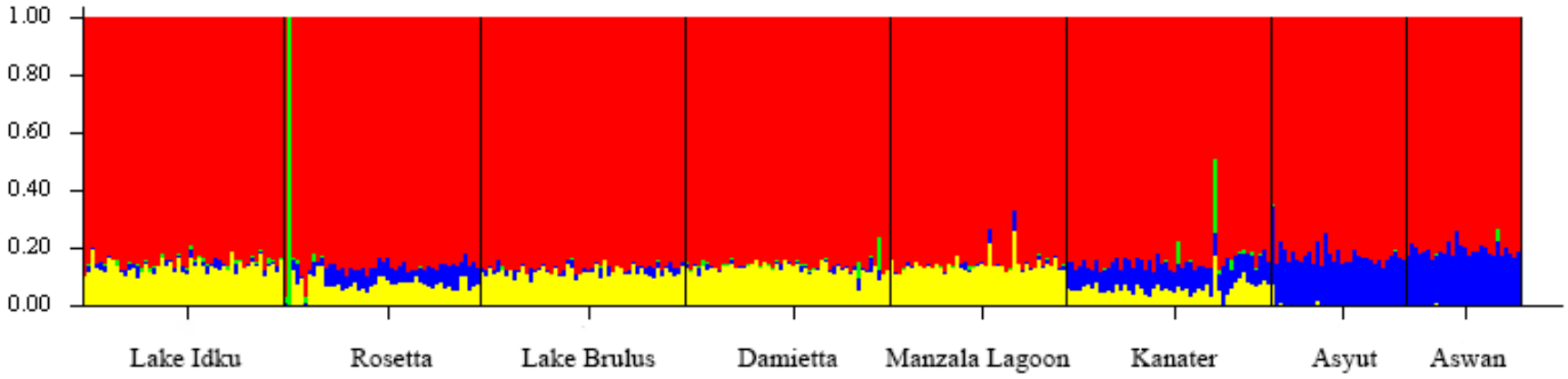
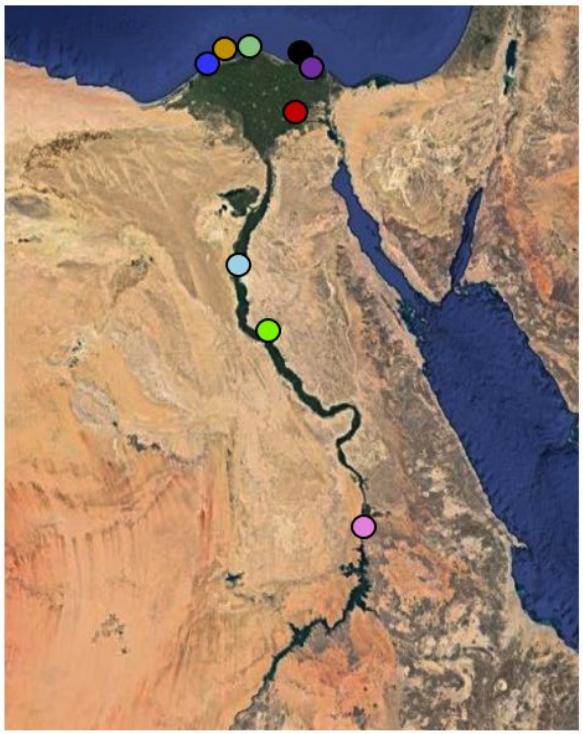


Fig. 4.22 Broad-Scale Population Structure. Structure plot of the eight wild sampling locations along a geographical gradient down the Nile River, Egypt at $\Delta K = 4$.

4.3.1.2 Fine-Scale Population Structuring

Mutual k-nearest neighbor analyses conducted in NetView pipeline v.1.1 to determine fine-scale population structuring exhibited a similar pattern to the STRUCTURE admixture analysis. The three generations of the AS form a distinct genetic cluster separate from the eight wild sampling locations, whilst the wild populations exhibit evidence of isolation-by-distance (Fig. 4.3; Fig. 4.4). The two most southern populations (Asyut and Aswan) are distinguishable from the populations further north and form a smaller, separate cluster (Fig. 4.4). However, a few individuals from these southern locations intermingled with northern samples indicated that there is still gene flow present between these populations (Fig. 4.3). There is a single sample from Damietta which clustered with the AS. Since Damietta is in close geographical proximity to the farm, it is conceivable that this individual is an escapee from the AS program (Fig. 4.3). Additionally, two individuals from the southernmost Aswan population formed a third clustering: indicating, that populations from further south in the Nile River and connecting waterways and lakes may exhibit greater variation amongst populations or include hybridized individuals (Fig. 4.3).

A.



B.

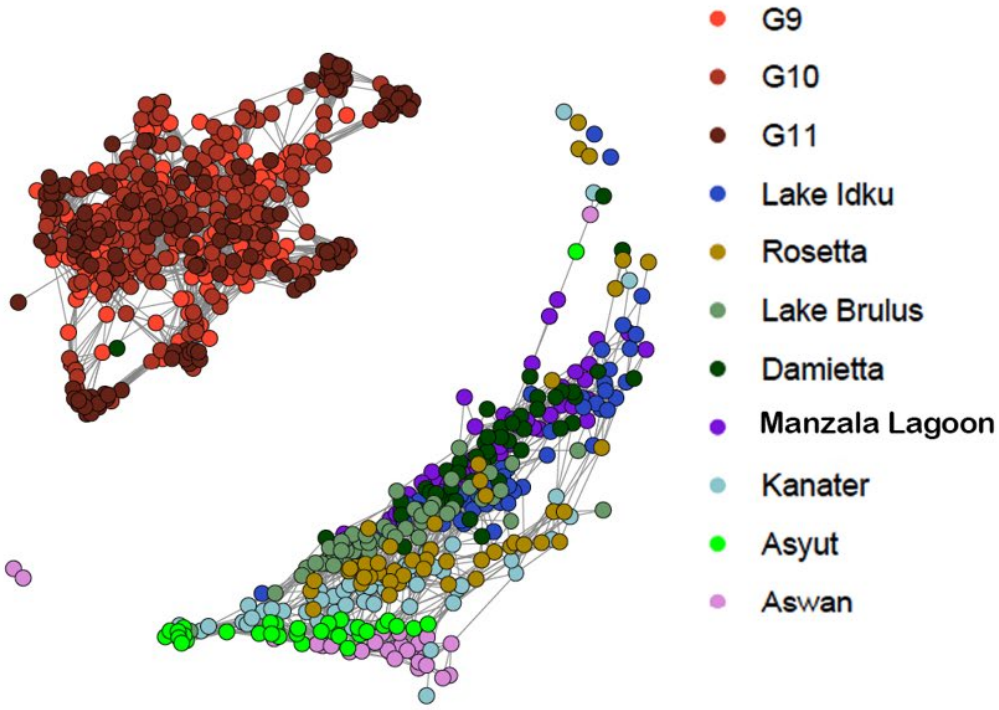


Fig. 4.23 Fine-Scale Population Structuring. A) Map of sampling locations along the Nile River in Egypt. B) Population clustering of all populations using an identity-by-state matrix constructed using the NETVIEW v1.1 pipeline at kNN = 20.

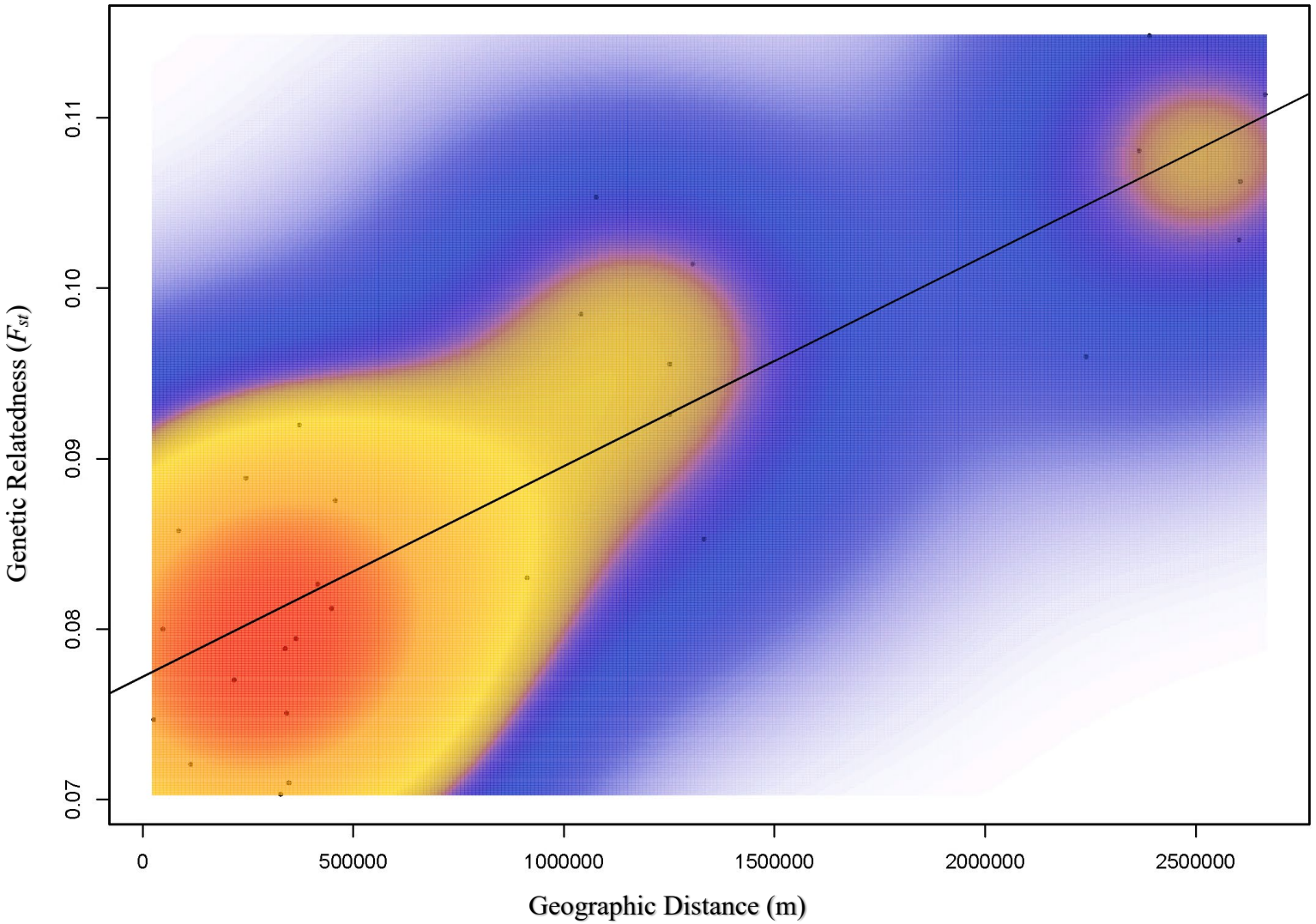


Fig. 24.4 Isolation-by-Distance. A heatmap comparing genetic relatedness (F_{st}) between two populations to their geographical distance (km) from one another. The heat map runs on a scale of red (higher relatedness) to blue (lower relatedness).

4.3.3 Signatures of Selection

QQ-plots examining the entire marker set revealed that the data violated the assumption of normality, indicating the presence of outliers (Appendix 13). A total of 674 outliers were jointly identified by both BayeScan and Arlequin between wild and domestic genetic clusters (Table 4.1). These outliers were confirmed by re-examining normality of the data using QQ-plots when the identified outliers were removed. QQ-plots revealed that the data conformed more to the assumption of normality than previously; however, there were likely still unidentified outliers in the dataset (Appendix 13). When all outliers identified by either BayeScan or Arlequin were removed from the dataset, they did not conform to the assumption of normality, indicating that those outliers identified by only one program were unlikely to be true outliers (Appendix 13). This confirmed the decision to utilize only jointly identified markers by both BayeScan and Arlequin when multiple sampling sites constituted a population (i.e. domestic or wild genetic clusters).

The greatest number of outliers (674) was found between the two genetic clusters identified using broad-scale population structuring analysis (Table 4.1; Fig. 4.1). Of those outliers 187 had negative alpha values in BayeScan and are under balancing selective forces, whereas the remaining 487 outliers had positive alpha values indicating directional selection. On average, pairwise comparisons of either Asyut or Lake Brulus to domestic populations (G9-11) yielded the greatest number of outliers (10-13; Table 4.1). The five wild populations which are most closely located in the Nile Delta (Rosetta, Lake Brulus, Damietta, Manzala Lagoon, and Kanater) had the fewest identified outliers (zero-three) when compared pairwise amongst themselves (Table 4.1). Regarding the pairwise comparisons of wild populations, Asyut vs. Rosetta had the greatest number of outliers (11) followed by Asyut vs Damietta (4; Table 4.1).

Outliers accounted for approximately 6.9% of the entire SNP marker set, with balancing outliers accounting for approximately 1.9% of the entire marker set and diversifying outliers accounting for approximately 5.0%. Diversifying outliers accounted for 72.3% of all identified outliers, whereas balancing outliers accounted for 27.7% (Digital Supplementary Material 2). Of the 674 identified outliers, 493 mapped back to the Orenil1.1 genome (Digital Supplementary Material 2). Every chromosome in *O. niloticus* had both directional and balancing outliers present, with the number of outliers per chromosome ranging from 9-61 (Digital Supplementary Material 2). A total of 193 outliers could be placed on the linkage map created in Chapter 3, with outliers mapping to five of the six QTL regions associated with sex and thirteen of the sixteen QTL regions associated with weight in at least one linkage map (Appendix 14).

Table 4.8 Pairwise Outlier Analysis All directional and balancing outlier loci identified. Outliers detected between the identified broadscale populations (i.e. Wild and Domestic) are those that were jointly identified by both BayeScan and Arlequin. Outliers detected between pairwise comparisons of sampled locations and AS generations are those detected by BayeScan, the more conservative of the two programs utilized for analysis.

	Gen 10	**Gen 11**	**Lake Idku**	**Rosetta**	**Lake Brulus**	**Damietta**	**Manzala Lagoon**	**Kanater**	**Asyut**	**Aswan**	**Wild**
Gen 9	0	11	5	3	11	4	6	6	13	6	
Gen 10		14	6	6	13	3	5	4	11	5	
Gen 11			2	4	13	5	5	4	10	1	
Lake Idku				0	0	0	0	0	1	0	
Rosetta					2	0	0	0	11	2	
Lake Brulus						0	3	1	0	1	
Damietta							1	0	4	0	
Manzala Lagoon								0	3	0	
Kanater									1	1	
Asyut										3	
Domestic											674

4.3.4 Population Genetic Diversity

The estimated effective population size for the AS ranged between 14.8-48.6 per generation, with only approximately 20-81% of each generation's breeding population genotyped (Table 4.2; Nayfa et al., 2020). Wild populations ranged from 30.5-infinite, with infinite being indicative of an infinite-sized ideal population and is taken to be an extremely high and positive value (Table 4.2; Jones et al., 2016). Despite these variations in effective population size, F_{is} values were all non-significant and negative in all AS generations and wild populations (Appendix 15). The proportion of SNPs that deviated from HWE in domestic populations were 2.8-14.6 times more frequent than in wild populations (Appendix 15).

Overall, the domestic population genetic cluster had higher expected heterozygosity (H_e), observed heterozygosity (H_o), multilocus heterozygosity *(*MLH), minor allele frequencies (MAF) and polymorphic loci than the wild genetic cluster when all, neutral, or directional markers were taken into consideration (Table 4.2). The greatest difference among populations and genetic clusters was observed when directional outlier markers were examined. When AS generations and wild population were individually considered, levels of both H_o and H_e for all and neutral markers were similar. In most instances, wild sampling locations (except H_o: Rosetta and Damietta and H_e: Rosetta, Damietta, and Kanater) had higher levels of heterozygosity than individual AS generations (Table 4.2). Rosetta had the lowest observed heterozygosity ($H_{o\_All} = 0.181$, $H_{o\_Neutral} = 0.180$. and $H_{o\_Directional} = 0.154$) and expected heterozygosity ($H_{e\_All} = 0.212$, $H_{e\_Neutral} = 0.210$, and $H_{e\_Directional} = 0.214$) in these three marker sets (Table 4.2).

The domestic populations, considered as a whole genetic cluster and individually, had a higher MLH overall than wild populations across three marker subsets (All, Neutral, and Directional; Table 4.2). Manzala Lagoon had the lowest MLH in all three marker sets ($MLH_{All} = 0.145$,

$MLH_{Neutral}$ = 0.130, and $MLH_{Directional}$ = 0.151; Table 4.1). However, when only balancing outlier markers were analyzed, genetic diversity indices for all populations and genetic clusters were similar to one another (Table 4.2).

The number of polymorphic loci per population ranged between 5,995-9,291 loci (61.0-94.5%), with domestic populations having 24.8% more polymorphic loci on average than the wild populations when all markers were considered (Table 4.2). A total of 565 private SNPs were identified within the domestic genetic cluster, while no private SNPs were identified within the wild genetic cluster.

As the number of polymorphic loci varied greatly between domestic and wild populations, the effect of missing data on genetic diversity indices was also examined (Appendix 16). Markers with less than 50%, 25%, and 5% missingness in all samples were tested as well as markers with a maximum of 5% missingness within a single population. As the percentage of missingness allowed per SNP decreased, the number of markers that passed this quality control measure also decreased. The number of polymorphic markers decreased from 61-95% when all markers were included to 44-67% when 50% missing data was allowed (Appendix 16). The percentage of polymorphic markers was similar between 25% missing data (29-44%) and 5% missing data per population (27-48%; Appendix 16).

In general, as the proportion of missing data allowed decreased, the number of polymorphic loci also decreased and estimates of observed and expected heterozygosity remained similar (±0.01) or decreased, with the exception of Rosetta at 25% missing data (Appendix 16). The marker set with only a total of 5% missingness per population allowed had the lowest number of polymorphic markers (8.5-16.4%), H_o, and H_e (Appendix 16). In population groupings with a larger number of individuals sampled (121-470 samples), heterozygosity estimates were less

affected and patterns remained more consistent than in groupings with fewer sampled individuals (28-50 samples; Appendix 16). Rosetta (48 samples), Asyut (33 samples), and Aswan (20 samples) showed the greatest variability among marker subsets (Table 4.2; Appendix 16).

Table 4.9 Genetic diversity indices calculated using all SNPs and subsets of SNPs (neutral markers, directional outlier markers, and balancing outlier markers). The estimated effective population size for each sampling location and/or timepoint calculated using linkage disequilibrium. Reported lower bound and upper bound numbers reflect a 95% confidence interval calculated using the jackknife method, a non-parametric method at a minimum allele frequency of 0.05. The average observed heterozygosity (H_o), expected heterozygosity (H_e), multilocus heterozygosity (MLH), minor allele frequency (MAF), and the number of polymorphic loci per sampling location and per STRUCTURE population designation (K = 2; domestic genetic cluster, wild genetic cluster).

	Category	n	LDN_e	$H_O \pm SE$				$H_E \pm SE$				MLH ± SE				MAF ± SE				Polymorphic Loci			
				All	Neutral	Directional	Balancing	All	Neutral	Directional	Balancing	All	Neutral	Directional	Balancing	All	Neutral	Directional	Balancing	All	Neutral	Directional	Balancing
Gen 9	Domestic	121	34.1-45.6	0.211 ± 0.013	0.207 ± 0.013	0.271 ± 0.013	0.284 ± 0.014	0.232 ± 0.014	0.226 ± 0.014	0.329 ± 0.012	0.291 ± 0.014	0.182 ± 0.002	0.178 ± 0.002	0.221 ± 0.005	0.271 ± 0.003	0.191 ± 0.006	0.126 ± 0.031	0.249 ± 0.014	0.203 ± 0.011	9,291	8,658	447	186
Gen10	Domestic	204	37.9-48.6	0.212 ± 0.010	0.206 ± 0.010	0.283 ± 0.010	0.280 ± 0.011	0.231 ± 0.011	0.224 ± 0.011	0.338 ± 0.009	0.290 ± 0.011	0.170 ± 0.002	0.166 ± 0.002	0.214 ± 0.003	0.255 ± 0.003	0.178 ± 0.004	0.127 ± 0.028	0.242 ± 0.014	0.206 ± 0.013	8,671	8,084	406	181
Gen 11	Domestic	145	14.8-21.9	0.211 ± 0.013	0.205 ± 0.012	0.283 ± 0.012	0.282 ± 0.013	0.229 ± 0.013	0.222 ± 0.013	0.339 ± 0.011	0.289 ± 0.012	0.175 ± 0.001	0.171 ± 0.001	0.222 ± 0.004	0.268 ± 0.003	0.185 ± 0.005	0.126 ± 0.022	0.247 ± 0.011	0.204 ± 0.010	8,934	8,338	414	185
Lake Idku	Natural	49	493.9-Infinite	0.214 ± 0.024	0.213 ± 0.023	0.200 ± 0.023	0.286 ± 0.024	0.232 ± 0.024	0.229 ± 0.024	0.247 ± 0.025	0.284 ± 0.022	0.133 ± 0.002	0.132 ± 0.002	0.102 ± 0.002	0.270 ± 0.005	0.156 ± 0.007	0.129 ± 0.022	0.153 ± 0.030	0.200 ± 0.020	6,404	5,970	251	183
Rosetta	Natural	48	30.5-Infinite	0.181 ± 0.029	0.180 ± 0.029	0.154 ± 0.029	0.274 ± 0.028	0.212 ± 0.029	0.210 ± 0.029	0.214 ± 0.031	0.286 ± 0.027	0.134 ± 0.002	0.133 ± 0.002	0.109 ± 0.002	0.261 ± 0.007	0.168 ± 0.009	0.125 ± 0.022	0.157 ± 0.028	0.199 ± 0.020	7,626	7,076	366	184
Lake Brulus	Natural	50	746.6-Infinite	0.221 ± 0.024	0.220 ± 0.024	0.208 ± 0.024	0.286 ± 0.023	0.236 ± 0.024	0.233 ± 0.024	0.251 ± 0.025	0.287 ± 0.022	0.142 ± 0.001	0.141 ± 0.001	0.110 ± 0.003	0.276 ± 0.004	0.169 ± 0.007	0.127 ± 0.022	0.144 ± 0.029	0.200 ± 0.020	6,754	6,285	285	184
Damietta	Natural	50	590-Infinite	0.202 ± 0.023	0.201 ± 0.023	0.180 ± 0.023	0.272 ± 0.023	0.225 ± 0.024	0.223 ± 0.024	0.226 ± 0.025	0.285 ± 0.022	0.136 ± 0.002	0.135 ± 0.002	0.107 ± 0.003	0.257 ± 0.005	0.166 ± 0.007	0.127 ± 0.024	0.149 ± 0.029	0.203 ± 0.021	7,041	6,551	307	183
Manzala Lagoon	Natural	43	175.4-Infinite	0.222 ± 0.025	0.221 ± 0.025	0.205 ± 0.025	0.271 ± 0.025	0.240 ± 0.025	0.238 ± 0.025	0.252 ± 0.026	0.284 ± 0.023	0.126 ± 0.004	0.125 ± 0.004	0.097 ± 0.004	0.247 ± 0.006	0.149 ± 0.007	0.130 ± 0.022	0.151 ± 0.033	0.200 ± 0.021	5,942	5,521	244	177
Kanater	Natural	50	297.3-Infinite	0.216 ± 0.025	0.214 ± 0.025	0.215 ± 0.027	0.305 ± 0.025	0.221 ± 0.024	0.218 ± 0.024	0.241 ± 0.026	0.290 ± 0.022	0.153 ± 0.006	0.152 ± 0.006	0.130 ± 0.007	0.292 ± 0.007	0.172 ± 0.008	0.152 ± 0.009	0.153 ± 0.028	0.205 ± 0.020	7,627	7,125	316	186
Asyut	Natural	33	149.3-Infinite	0.268 ± 0.042	0.266 ± 0.042	0.270 ± 0.042	0.314 ± 0.044	0.259 ± 0.036	0.257 ± 0.036	0.292 ± 0.037	0.287 ± 0.034	0.164 ± 0.005	0.163 ± 0.005	0.133 ± 0.006	0.301 ± 0.008	0.182 ± 0.013	0.149 ± 0.012	0.147 ± 0.036	0.205 ± 0.026	6,553	6,106	260	187
Aswan	Natural	28	55.7-Infinite	0.247 ± 0.032	0.245 ± 0.042	0.251 ± 0.034	0.290 ± 0.033	0.265 ± 0.031	0.264 ± 0.031	0.288 ± 0.031	0.290 ± 0.029	0.137 ± 0.003	0.136 ± 0.003	0.107 ± 0.004	0.277 ± 0.006	0.169 ± 0.009	0.152 ± 0.012	0.147 ± 0.041	0.203	5,995	5,581	228	186
Domestic		**470**		**0.208 ± 0.007**	**0.203 ± 0.007**	**0.271 ± 0.006**	**0.281 ± 0.007**	**0.228 ± 0.007**	**0.222 ± 0.007**	**0.330 ± 0.006**	**0.289 ± 0.007**	**0.175 ± 0.001**	**0.171 ± 0.001**	**0.218 ± 0.002**	**0.263 ± 0.002**	**0.184 ± 0.003**	**0.151 ± 0.006**	**0.246 ± 0.007**	**0.204 ± 0.006**	**9,234**	**8,609**	**439**	**186**
Natural		**351**		**0.177 ± 0.009**	**0.176 ± 0.009**	**0.147 ± 0.009**	**0.286 ± 0.008**	**0.165 ± 0.009**	**0.191 ± 0.009**	**0.185 ± 0.010**	**0.285 ± 0.008**	**0.035 ± 0.001**	**0.140 ± 0.001**	**0.111 ± 0.002**	**0.272 ± 0.002**	**0.166 ± 0.003**	**0.127 ± 0.008**	**0.150 ± 0.011**	**0.202 ± 0.008**	**8,577**	**7,979**	**412**	**186**

4.4 Discussion

This study used genome-wide SNP markers to 1) investigate population genetic structure 2) detect signatures of selection in three generations of the AS and eight wild populations of Nile tilapia (*O. niloticus*; Aswan, Manzala Lagoon, Kanater, Lake Idku, Damietta, Lake Brulus, Rosetta, and Asyut) throughout the Nile River, Egypt, and 3) audit genetic diversity in the AS and wild populations.

Clear population genetic structuring was observed indicating that the domesticated AS genetic cluster has become genetically distinct from the wild genetic cluster in Egypt. The genetic distinction between the AS and wild populations is likely due to the initial bottleneck created by a small founding population, genetic drift and the subsequent selection for faster growth rates, larger sizes, and domestication within this limited population. This clear separation between wild and domestic populations has also been observed in Atlantic Salmon, *Salmo salar* (Gutierrez et al., 2016) and gilthead sea bream, *Sparus aurata* (Cossu et al., 2019). The effects of the bottleneck created by the small founding population for AS can be observed in the smaller effective population size (max 48.6) of the domesticated AS in comparison to the wild effective population size (max 'infinite'). Similar results have been seen in other aquaculture species, like Atlantic Salmon, *Salmo salar* (Domestic N_e 33-125, Wild N_e = 50- >20,000; Bentsen and Thodesen, 2005), Pacific oyster, *Crassostrea gigas* (Domestic N_e = 47.6-58.5, Wild N_e= 527.9-infinite; Zhong et al., 2017), and gilthead sea bream, *Sparus aurata* (Domestic N_e = 21-111, Wild N_e = 133-infinity with the exception of one domestic population; Cossu et al., 2019).

Despite evidence of gene flow among the eight wild populations, isolation-by-distance was detected with the two most southern populations (Asyut and Aswan) being more distinct from the Nile Delta populations to the north than the geographically intermediate Kanater population.

In addition to the effects of physical distance to gene flow and population structure, environmental factors may have also influenced this distinction between Delta and upstream populations. Individuals within Delta populations, particularly Lake Idku, Lake Brulus, and Manzala Lagoon, which have a direct connection to the sea, live in brackish to freshwater conditions whilst the individuals within the upstream populations live in freshwater conditions (Balah, 2012, Hassanien et al., 2004).

These results are similar to those observed in 2004 and 2005 in two separate studies using microsatellites and randomly amplified polymorphic DNA (RAPD) where evidence of population sub-structuring was identified. Structuring in these studies was not only identified between geographically distant Nile Delta populations and upstream Egyptian Nile populations, but also amongst lake and river base populations in the Delta (Hassanien and Gilbey, 2005, Hassanien et al., 2004). However, unlike those studies, the present study observed no significant population structuring among Nile Delta populations. This disparity may be attributed to the difference in molecular technologies utilized between studies and the dramatic rise in aquaculture in Egypt (Soliman and Yacout, 2016).

Differences in molecular technologies have likely contributed to the disparities in population structure. For instance, Hassanien and Gilbey (2005) inferred the presence of null alleles based on lower levels of observed vs. expected heterozygosity levels in their microsatellite dataset. Null alleles in microsatellite studies can result in the overestimation of F_{st} and genetic distance (Chapuis and Estoup, 2006). Whereas, the RAPDs used in Hassanien et al. (2004) are limited by the fact that the majority of RAPD markers are dominant, making it impossible to determine whether a DNA segment is amplified from a homozygous or heterozygous locus (Kumar and Gurusubramanian, 2011). This can result in uncertain estimates to genetic structure (Fritsch and

Rieseberg, 1996). Additionally, the molecular criteria which determine what constitutes population structure are flexible and can vary based on the organism, study question, and genetic markers used (Putman and Carbone, 2014, Waples and Gaggiotti, 2006).

Genetic technologies are not the only factor to have changed over the years. Since 2005, Egypt has experienced a considerable increase in extensive, semi-intensive, and intensive farming systems for Nile tilapia (Soliman and Yacout, 2016).The vast majority of these farms are located in the Nile Delta region and concentrated in the Northern Lakes (Maruit, Idku, Brulus, and Manzala Lagoon; Soliman and Yacout, 2016). As a result, increased movement of fish among hatcheries and farms has occurred in the region in that time. In addition, the number of fish escaping from farms has likely increased due to a combination of local weather conditions, including flash flooding events (Moawad et al., 2016), and farm practices. With five of the eight sampled locations in the Nile Delta regions, and farming occurring at or near the remaining three sampling locations (Soliman and Yacout, 2016), the genetic diversity of the wild populations may have been affected by exchange with farmed stocks.

A comparison of wild and domestic genetic clusters identified 674 outlier markers, with a higher proportion of markers deviating from HWE in domestic populations than wild populations. This is indicative of a finite population size and selective forces, such as artificial selection for marketable traits and domestication (Waples, 2014). The large amount of outliers detected concurs with other genetic studies of domestic vs. wild aquatic populations, including brown trout *Salmo trutta* L., (431 SNP outliers; Linløkken et al., 2017) and Atlantic salmon, *Salmo salar* L. (337 and 270 SNP outliers; López et al., 2019). Both balancing and diversifying outliers identified between domestic and native populations were found in every chromosome. Unlike other studies which found specific regions of the genome under selection when comparing

domestic and wild populations (López et al., 2019, Marrano et al., 2018), there was a lack of localized clustering of outliers. However, outliers were detected on regions of the genome associated with sex and weight, supporting that these traits are polygenic traits influenced by genes multiple chromosomes (Chapter 3).

A limited number of outliers (0-11) detected in pairwise comparisons of wild populations is consistent with the limited genetic differentiation observed among the wild populations. The fact that the number of outliers detected increased with geographic distance from the upstream (Asyut, Kanater, and Aswan) to Nile Delta populations (Lake Idku, Rosetta, Lake Brulus, Damietta, and Manzala Lagoon) also reflects the isolation-by-distance determined using the whole data set. These results suggest that despite known differences in salinity levels in delta and upstream populations, there appears to be little or no effect on selection. This is not entirely surprising as Nile tilapia are known for their tolerance to a wide range of environmental conditions (Avella et al., 1993, Balarin and Haller, 1982, Chervinski, 1982, Dominguez et al., 2004, Kamal and Mair, 2005, Philippart and Ruwet, 1982, Randall and Tsui, 2002, Rebouças et al., 2016, Shelton and Popma, 2006). Alternatively, gene flow may be high enough between geographic regions to combat the forces of natural selection (Lenormand, 2002). Consequently, few outliers amongst wild populations indicate that the AS would be expected to perform similarity in different locations once disseminated throughout Egypt.

Differences in genetic diversity resulted in the domesticated AS being clearly distinguishable from wild populations. In general, genetic diversity indices indicate that AS populations have higher levels of heterozygosity than wild populations. This held true regardless of the number of SNPs and levels of missing data allowed. These results differ from what is traditionally seen in domesticated and/or selectively bred populations vs. wild populations where wild populations

exhibit either higher levels of genetic diversity (Makino et al., 2018, Zamani et al., 2018), or similar levels of heterozygosity (Gutierrez et al., 2016). This may be explained by 1) hybridization with another tilapia species 2) the isolation-by-distance observed in this study among current wild populations and 3) the historical development of fishing and aquaculture in Egypt.

The AS had a higher number of polymorphic markers and private SNPs (5.7 % of all SNPs) than wild populations. While this may be a result of domestication or founder effects, it is suspected that introgression has occurred with blue tilapia (*Oreochromis aureus*). This interpretation is further supported by the large number of outliers detected, as hybridization has been interpreted to explain the detection of outliers in other species (Cullingham et al., 2014). Blue tilapia from a population maintained at the Abbassa Station, Egypt have been observed in earthen ponds in AS facilities (Benzie, 2019; pers. comm.). This population of blue tilapia has now been removed from the Abbassa Station. Unpublished research by the WorldFish Center and affiliated researchers has found that the AS is comprised of 10% *Oreochromis aureus* (blue tilapia; Grobler, 2017). Species-specific SNPs are often picked up when developing SNPs from samples that include multiple species or hybrids (Liu et al., 2011, Silva-Junior et al., 2015). Thus, the incorporation of *O. aureus* in the AS genome may account for the high number of private SNPs identified in the AS, as well as the higher number of polymorphic markers and heterozygosity observed in the AS genetic cluster over the wild genetic cluster as these markers may have been species-specific SNPs. While the AS showed the greatest number of polymorphic loci, the wild populations all exhibited different subsets of polymorphic loci per sampling location, indicating that hybridization with *O. aureus* may have also occurred in the wild. Given that tilapia species

are well known for hybridizing in both aquaculture and wild environments, this is unsurprising (D'Amato et al., 2007, Deines et al., 2014, Lovshin, 1982, Meier et al., 2019).

Considering the level of genetic distinction between wild and domestic populations of Nile tilapia described within this research, putative AS escapees were easily identified, with suggested evidence of first and later generation escapees in Rosetta, Kanater, and Damietta detected. Escapees in other locales, particularly from selectively bred individuals, have been shown to lower the fitness of wild populations (Yang et al., 2019) as demonstrated in Atlantic salmon, *Salmo salar* (Glover et al., 2013, McGinnity et al., 2003); European seabass, *Dicentrarchus labrax* (Toledo-Guedes et al., 2014); and Turbot, *Scophthalmus maximus* (Prado et al., 2018). It is not clear to what extent this may be a concern for tilapia, because while there was evidence in the present study of AS genetic material in wild Egyptian populations, there is at present, no information on fitness differentials between domesticated and wild tilapia populations.

High levels of genetic diversity were still observed within the AS, suggesting that the potential detrimental effects on diversity of any AS escapees that do survive in wild populations may be minimal. This is particularly true as the AS was founded from both Nile Delta and upstream populations of Nile tilapia in Egypt. Thus, the genetic diversity observed in the AS is a subset of what is already available in wild populations. This in addition to the relatively low number of escapees detected when considering all wild populations, suggests that escapees may either be a rare occurrence or may have low survival within wild populations. This has been demonstrated previously in domestic rainbow trout (*Oncorhynchus mykiss*) who experience lower survival rates in the wild due to their increased size and bolder foraging habits exposing them to higher predation (Biro et al., 2004). Regardless, continued monitoring of escapees from the AS and other domestic lines is important as many wild Nile tilapia populations are at risk of an altered

population structure and genetic diversity due to anthropogenic changes such as habitat disturbance, overfishing, and indiscriminate fish transfers of tilapia species throughout Africa (Eknath and Hulata, 2009).

4.5 Conclusions

The present study has highlighted the valuable information for improved management of aquaculture species by investigating population genetic structure, genetic diversity, and signatures of selection between domestic and wild populations. In the case of Nile tilapia in Egypt domestic and wild populations were found easily distinguishable from one another using SNP markers, even when compared to founding populations. In turn, this distinct clustering allowed for easy detection of putative escapees. Although the wild genetic cluster was not panmictic, with wild populations displaying evidence of isolation-by-distance, levels of genetic differentiation were relatively low and no evidence of significant signatures of selection among wild populations were observed. Despite 11 years of selective breeding, the AS displayed high levels of genetic diversity. These data suggest that the AS could be disseminated throughout Egypt with negligible differences in performance expected and minimal disruption to wild populations. The genetic diversity comparisons also helped better understand how the effects of selection, founder effect, inbreeding, and genetic drift have affected this domestic line. The effects of substantial pedigree errors may have slowed selection for growth as well as the introgression with *O. aureus.* This introgression may also explain the large number of outliers detected between wild and captive genetic clusters. While both balancing and diversifying outliers were traced back to all 22 *O. niloticus* chromosomes, additional research is required to determine the nature of these signatures and their direct relevance to biological or evolutionary processes within domestic and wild populations.

CHAPTER 5: GENERAL DISCUSSION

5.1 Significant Findings

In recent years, the number of aquatic selective breeding programs have increased, with most programs using a closed nucleus mating system where no new genetic material is introduced within the line (FAO, 2019a, Gjerde, 2005). Consequently, monitoring the genetic status of these selective breeding programs to reduce loss of genetic diversity while obtaining genetic gain is vital. The Abbassa Strain (AS) was first developed by WorldFish as a selective breeding program to improve production of Nile tilapia in Egypt (Ibrahim et al., 2013, Rezk et al., 2009). Prior to the investigations described in Chapters 2-4 in this thesis, little was known about the genetic state of the AS: including, the accuracy of pedigree traceability and maintenance of genetic diversity within the selected line, understanding commercially important traits, identifying signatures of selection, the extent of wild population structuring of Nile tilapia in Egypt, and finally, the potential effects AS escapees may have within the regions it is currently farmed, and also across Egypt where it is intended to be disseminated.

To date, management of the AS has been entirely reliant on genealogical data; however, this study has shown that genealogical records in the AS are 3-50 times more erroneous than reported for terrestrial genetic improvement programs (Chapter 2). Molecular data was used to correct these records, with reassignment rates increasing as the percentage of broodstock genotyped increased. Of these reassignments, 23.9% were family-based errors, where offspring were reassigned to a sibling of their originally recorded parent, and the remaining 76.1% have been categorized as “random” with the available data from farm management. Despite this high error rate, genetic diversity has been maintained and is comparable to, and in some cases higher than,

genetic diversity within wild populations (Chapter 2, Chapter 4). Inbreeding levels have remained well below the acceptable 1% increase per generation for a selective breeding program (Chapter 2; Bentsen and Olesen, 2002, Woolliams, 1994); nevertheless, founder contribution has degraded, with only 34 of the original 83 founders accounting for 84.3% of the genetic variability currently within the AS.

To understand the genetic architecture of weight and sex, quantitative trait locus (QTL) analyses and genome-wide association studies (GWAS) were conducted within the AS. Due to the high error rates in genealogical records, originally established mapping families were not viable for traditional family-based mapping methods, therefore a novel population-based linkage mapping method was used, with independent validation based on genome assemblies for *O. niloticus* (Chapter 3). Due to power constraints within this method, unique linkage maps were established for the sex average, female, and male lines. A total of 2,399 markers were successfully mapped to their most likely positions within 21 of 22 linkage groups for the sex average map, whereas 2,197 markers mapped to the female-specific map and 2,125 markers to the male-specific map. This study then identified multiple suggestive QTL and genetic associations for both sex and weight, adding support to the notion that these two traits are polygenic in the AS.

Additionally, outlier analyses were conducted within the AS line and across eight wild populations of Nile tilapia in Egypt with 6.9% (674) of all markers categorized as outliers between the two identified genetic clusters (wild and domestic; Chapter 4). Nonetheless, interpretation of these signals of selection was made more complex as the AS has undergone introgression with wild Blue tilapia (*Oreochromis aureus*), with an estimated 10% of its genome attributed to *O. aureus* and 90% to *O. niloticus* (Chapter 3; Grobler, 2017). This introgression

within the AS was further supported by the detection of 565 private alleles within the AS despite them being entirely sourced from Egyptian populations (Chapter 3, Chapter 4).

While the AS was the only population to exhibit private alleles, the wild populations all exhibited different subsets of polymorphic loci per sampling location, indicating that hybridization with *O. aureus* may have also occurred in the wild. Within the wild cluster some isolation-by-distance was observed between brackish Nile Delta populations and upstream riverine populations. However, only a limited number of outliers (0-11) were detected in pairwise comparisons of wild populations, suggesting that environmental selection is not driving the observed isolation-by-distance, or that gene flow among populations is sufficient to counter selection (Lenormand, 2002). A putative first-generation AS escapee was detected in the wild, demonstrating that individuals from the AS are easily distinguishable from their wild counterparts and that their effects on the populations can be monitored (Chapter 4).

5.2 Determining Management Effectiveness for a Selective Breeding Program

Management effectiveness within a selective breeding program can be measured by the rate of pedigree errors in genealogical records. Pedigree error rates within the AS were demonstrated to be between 45-51% per generation (Chapter 2). To this author's knowledge, error rates in genealogical records in other aquatic breeding programs have remained proprietary information and have been rarely reported in the literature with only a single conference presentation on a selective breeding program for *Litopenaeus vannamei* (16.2% maternal errors and 21.2% paternal errors detected in pedigree records) available online (Jerry et al., 2017). However, error rates previously detected in the pedigrees of most non-aquatic breeding programs have been in a range of 1-15% (Bovenhuis and Van Arendonk, 1991, Crawford et al., 1993, Doerksen and Herbinger, 2010, Sanders et al., 2006). Whether these ranges also apply to a standard aquatic

selective breeding programs is yet to be determined, as the life histories of these animals including, high fecundity (Gjedrem and Baranski, 2010c), larval sizes below minimum sizes for physically tagging individuals (Ouedraogo et al., 2014), and asynchronous spawning (Nguyen, 2016) may make these programs more susceptible to pedigree errors. However, these errors can be minimized through molecular data and stricter farm management and recording measures.

It is essential that pedigree errors within a selective breeding program be reduced, as these errors lead to incorrect estimates of additive variance, decreased prediction accuracy of estimated breeding values (EBVs), reduced accuracy of genomic selection predictions, and ultimately diminished genetic gain (Munoz et al., 2014, Nwogwugwu et al., 2019). Evaluations of the effects of pedigree errors on selective breeding programs found that a 10% pedigree error rate resulted in genetic gain being 2.0-4.3% lower than if correct records had been utilized (Israel and Weller, 2000, Visscher et al., 2002). If error rates in aquatic selective breeding programs are higher than 10%, an even greater reduction in genetic gain can be expected, and the targeted 10% increase in productivity per generation of a well-designed and maintained aquatic selective breeding program will not be achieved (FAO, 2019a). Thus, improving pedigree traceability in these programs is imperative to increase program efficiency and improve global food production at a rate that can meet the rising demand.

Although the direct effect of pedigree errors on the AS were not quantified in this study, the level of genetic gain observed (3.8-7.0% genetic improvement per generation; Rezk et al., 2009) was approximately half that of the GIFT strain of Nile tilapia (7.1-15.0% genetic improvement per generation; Eknath and Acosta, 1998, Ponzoni et al., 2011). As the two programs were derived from *O. niloticus* sourced from Egyptian populations (Eknath et al., 1993, Ibrahim et al., 2013, Rezk et al., 2009, WorldFish, 2016), and all wild Egyptian populations have been shown to share

similar levels of genetic diversity (Chapter 4), pedigree errors have likely contributed to a reduction in genetic gain within the AS. However, as these two strains were established in different environments, GxE interactions may have also had an effect on the observed disparity in genetic gain. GxE studies of both the AS and the GIFT strain found little to zero GxE interactions within the same country (Eknath et al., 1993, Khaw et al., 2009); however, significant GxE interactions have been observed in the GIFT strain when environmental conditions in different countries were examined (Agha et al., 2018). As such, if there is a substantial disparity in genetic gain between programs of the same species, as that observed between the AS and GIFT strains, it is recommended that GxE interactions be examined to determine the extent of environmental conditions influencing strain performance.

Comparisons with wild populations revealed 565 private alleles within the AS (Chapter 4). This was initially unexpected given that the AS was derived entirely from native Egyptian populations and wild samples had been collected from the entire length of the Nile River in Egypt for comparison. Although wild populations showed evidence of isolation-by-distance, genetic diversity within these populations was similar to one another (Chapter 4). Another study of the AS found that approximately 10% of the AS genome was derived from *O. aureus* in addition to approximately 90% *O. niloticus* (Grobler, 2017). As the incorporation of *O. aureus* genetic material was not an objective for the AS, the inclusion of *O. aureus* in the AS resulted from unintended hybridization. This hybridization would have occurred between the AS and an *O. aureus* population derived from wild Egyptian stock also maintained at the Abbassa Station until recently (Benzie, 2019; pers. comm.). Flooding events had been recorded where there was some mixing of fish on the site and specific records of *O. aureus* individuals being photographed from the pond containing the AS rearing hapas. Therefore, there is reasonable evidence for the

possibility of hybridization occurring at the site possibly after the initial development of the AS genetic improvement program and from lack of adequate management of the AS population. Additionally, as the private SNPs were observed throughout the AS, it suggests that *O. aureus* genetic material is well integrated within the entire selective breeding program indicating that this hybridization occurred early in the program and may also be the result of several separate hybridization events.

Interspecific hybridization has been observed more often in fish, than in any other vertebrate group (Scribner et al., 2000) due to a number of factors including, external fertilization and a scarcity of conspecific mates (Hubbs, 1955, Willis, 2013). Inadvertent hybridization and introgression with a parent species is not a new phenomenon in aquaculture, particularly with species, such as carp, the most farmed fish globally (FAO, 2018), that are known to readily hybridize in wild settings as well as domestic (Mia et al., 2005, Padhi and Mandal, 1997). This incidental hybridization and introgression with the selective breeding line, can result in unexpected and undesirable results in hybrid progeny, including genetic deterioration and changes in the expression of desirable traits (Rahman et al., 2013). Thus, the inclusion of *O. aureus* may have also had an effect on the modest genetic gain observed within the AS. Another study examined the growth rates of purebred *O. aureus, O. niloticus,* and their F1 hybrids, finding that purebred *O. aureus* do not grow as large as *O. niloticus* while their hybrids exhibited faster growth rates than either species (El-Hawarry, 2012). No subsequent studies have been conducted on backcrossing F1 hybrids of *O. aureus* and *O. niloticus* with *O. niloticus*; however, some investigations into cichlid hybrids have been reported. Hybrids of seven cichlid fish species, were found to exhibit hybrid vigor in the F1 generation, but the effect was lost in subsequent generations when recessive allelic incompatibilities surfaced (Dobzhansky, 1936,

Stelkens et al., 2015). As the hybrid vigor initially experienced by other cichlids was lost in subsequent generations, this suggests that we can expect a similar result from *O. niloticus* and *O. aureus* hybrids; however, hybridization of these two species and introgression with *O. niloticus* should be further examined to better understand the impact this has had on the AS.

Overall, this thesis has shown that molecular data can be used to determine management effectiveness of selective breeding program. Pedigree error rates within the AS exceeded those observed in terrestrial breeding programs by 3 to 50 times and evidence of an unintentional hybridization with *O. aureus*, suggests that the AS has not been managed properly. Given the differences in life histories between aquatic and terrestrial species, the risk of accumulating pedigree errors from incorrect records and unexpected hybridization is higher in aquatic selective breeding programs than in terrestrial programs. To counteract this, aquatic programs must be more diligent in their records and would benefit from incorporating molecular techniques to check and correct records as needed. In programs, like the AS, where pedigree errors are high and have accumulated in the program for an undetermined number of years, the genetic status of the program should be examined to determine the viability of salvaging the program and measures should be taken to restructure current procedures on farm and educate staff on the proper management of a selective breeding program.

5.3 Understanding the Current and Ongoing Genetic Status of a Selective Breeding Program

In an effort to maximize the retention of genetic diversity within a selective breeding program, pedigree records are used to monitor family lines and ensure that an adequate number of families are maintained per generation (Skaarud et al., 2014). Despite maintaining 100-120 families per generation in the AS, only 41% of the original founders account for the majority of the AS's

genetic composition (84.3%; Nayfa et al., 2020), indicating that tracing family lines alone in this case is not adequate to maintain genetic diversity within the selective breeding program. This is due to the nature of selection. Selection of individuals from each family is not random, and as selective breeding programs progress, high performing individuals selected for broodstock likely exhibit a higher proportion of genetic material from the same few high performing founders (Sonesson et al., 2012). This effect has likely been somewhat mitigated in the AS due to pedigree errors resulting in a partial random mating scheme. Thus, retention of founder genomes may have been further reduced had the program not experienced these errors. Founder contribution can still be improved in a long-term selective breeding program like the AS. For example, research into other aquatic species, like red sea bream, has demonstrated that "lost" genetic variation can be recovered within a line with over- and under- represented founders, by selecting broodstock which maximized the number of founder lineages present (Doyle et al., 2001).

Although founder contribution was not well maintained in the AS, genetic diversity was found to rival that of wild populations and inbreeding was determined to be within an acceptable level for the age of the program, both likely due to the partial random mating scheme and unintentional hybridization with *O. aureus*. Thus, it is suggested that the AS selective breeding program is salvageable. In order to maintain genetic diversity and reduce inbreeding moving forward with the AS, and in similar selective breeding programs, measures that minimize pairing of related individuals should be fully incorporated. These pairings should also consider the effective number of founders within the line, as they represent the full genetic diversity available within a closed system breeding program.

Another option to improve genetic diversity in a selectively bred line is through the incorporation of new germplasm. While cross breeding different domesticated lines has been proven to

improve genetic diversity, reduce inbreeding, trigger heterosis and subsequently improve animal performance (Goyard et al., 2008, Stronen et al., 2017), the long term effects (i.e. beyond the F1 generation) of these inclusions have rarely been examined. The present study observed that the inclusion of new germplasm, as seen in Generation 4 of the AS, and the genetic benefits of these new individuals can be short-lived if they are not actively selected for in future generations, or if new germplasm isn't incorporated into the line in frequent intervals (Chapter 2). As such, selective breeding programs should employ similar measures to incorporate secondary founders as those described previously for retaining and reestablishing founder genomes within a closed system breeding program.

5.4 Developing Genetic Tools for a Selective Breeding Program

In order to understand the genomic architecture of commercially important traits, it is useful to establish a line specific linkage map. Linkage mapping relies upon recombination rates to calculate the relative position of markers to one another; however, recombination rates can vary between species, populations, individuals, and even genomic regions (Dukić et al., 2016). Maps of swordtail species (Schumer et al., 2014, Schumer et al., 2018) and cichlid species (Bezault et al., 2012, Meier et al., 2017) have detected different recombination rates and segregation distortions between hybrid and purebred parents (Payseur and Rieseberg, 2016), further supporting the need for strain specific maps for lines that have experienced hybridization and/or introgression, like the AS.

High levels of pedigree errors identified in the original mapping families (Chapter 2) led to the utilisation of a population-based method for linkage mapping (Chapter 3). This method incorporated the recombination frequencies calculated per family, based on maternal lines, paternal lines and the sex average maps, into a weighted population average for recombination

frequencies per SNP. Although this method has been developed in linkage mapping programs like JoinMap v. 5 (Stam, 1993, Van Ooijen, 2018), it is not widely used and has not been validated in the general literature. However, given the genomic resources already available for Nile tilapia, a model species for aquaculture, the reduced mapping family dataset for the AS provided the unique opportunity to not only create, but also validate the population-based method for linkage mapping. Within comparisons between *de novo* and genome-based mapping order to Orenil1.1 (GenBank Assembly Accession: GCA_00188235.2), 98.6-99.7% of markers agreed in linkage group placement. Validating a method that only requires a small number of individuals per family is particularly important for non-model species, markedly where fewer offspring are available, or in cases where sampling is not ideal as it denotes that linkage mapping can be conducted even if family numbers and sizes do not meet the requirements for more traditional linkage mapping approaches. Nevertheless, to build further confidence in this novel methodology for linkage mapping, it is recommended that it be validated using other species.

Linkage mapping using *de novo* genotype by sequencing SNP discovery methods can restrict reproducibility and comparisons to other maps since marker sets can change with each run. However, the need to create independent maps for each sampling subset can be overcome with the use of a SNP array, such as a solid-state probe hybridization array [i.e. Affymetrix| ThermoFisher Scientific and Illumina]. A SNP array would not only allow for the calling of the same markers for each genotyped sample, but increases genotype accuracy for a much larger number of markers and allows for greater control in selecting a more even marker distribution across the genome (Robledo et al., 2018). The development and use of SNP arrays for Nile tilapia are becoming more accessible with one SNP array already developed for *O. niloticus* (Joshi et al., 2018) and another *O. niloticus* array incorporating all WorldFish strains in

development (Benzie, 2019, pers. comm.). Once these arrays are utilized in specific strains, maps can be created and refined over time, allowing for the pool of genotyped samples to grow, lessening the financial burden on selective breeding programs and increasing their ability to pursue GS. Additionally, if the same arrays were used across strains, it would allow for the direct comparison of these strains to one another to identify differences in selection, which may be useful in understanding differences in strain performance.

5.5 Understanding the Genetic Architecture of Commercially Important Traits

In order to pursue MAS or GS to improve production within a selective breeding program, it is vital to have an understanding of the genetic architecture of commercially important traits, such as sex and weight (Zenger et al., 2019). This can be accomplished through QTL mapping and GWAS approaches as these analyses identify the genomic regions involved in a trait and can be used to determine their effect sizes. These in turn can be incorporated into genomic estimated breeding values (GEBVs) to improve the accuracy of existing EBVs or phenotypic based breeding programs (Lee et al., 2015, Zenger et al., 2019) and pursue GS (Hayes and Goddard, 2010, Zenger et al., 2019).

Although QTL detection through QTL mapping and GWAS approaches traditionally utilize a larger number of samples, the present study demonstrated that both approaches could be conducted despite small families sizes (≤ 10 offspring) and a relatively small sample set with phenotypic data available (388 individuals). QTL mapping approaches can be quite powerful in identifying associations between genotypes and phenotypes, particularly when families with a large number of offspring are available (Korte and Farlow, 2013), whereas genome-wide association studies (GWAS) rely on large heterogeneous populations to unravel linkage disequilibrium in markers (Hayes, 2013, Korte and Farlow, 2013). By these definitions, GWAS

is the more powerful QTL detection method for population-based linkage mapping. However, in this study, GWAS only detected QTLs for sex determination in LG23; whereas, QTL mapping analysis identified a number of QTLs associated with both weight and sex. However, as QTL mapping families were small and some QTLs varied per family, independent validation would be required to understand their relevance to the wider AS.

Putative QTLs, or QTLs identified in both families, were identified for sex, but not for weight in the AS. Sex determination is notoriously complex for Nile tilapia, with sex chromosomes still unclear despite the numerous linkage maps and two genome assemblies available (Cáceres et al., 2019, Conte et al., 2017, Eshel et al., 2011, Grobler, 2017, Joshi et al., 2018, Lee et al., 2003, Lin et al., 2016, Liu et al., 2014, Palaiokostas et al., 2015). Leading evidence suggests that *O. niloticus* exhibits a male heterogametic (XX|XY) sex determining system (Mair et al., 1991); however, numerous other genomic regions have been identified as being sex-linked and in some cases sex-determining (Cáceres et al., 2019, Conte et al., 2017, Eshel et al., 2011, Lee et al., 2003, Palaiokostas et al., 2015). In this case, the AS has also been shown to have hybridized with *O. aureus* (Grobler, 2017), which is thought to exhibit a female heterogametic (ZW|ZZ) sex determining system (Campos-Ramos et al., 2001), further complicating the detection of genomic regions associated with sex for this strain. This hybridization has resulted in the detection of novel regions of the genome segregating with sex as described in Chapter 3. To understand the full extent of the effects hybridization has on sex-associated regions more comprehensive and targeted studies are needed to associate these QTLs segregating with sex to sex determination in *O. niloticus* which has experienced introgression with *O. aureus*.

Although no GWAS or putative QTLs were identified for weight, suggestive QTLs, or QTLs identified in a single family, were detected in 16 linkage groups indicating that there is still

significant family specific variation for weight within the AS. This is supported by data from other studies of Nile tilapia as well as in other tilapia species which have all indicated that growth rate is a polygenic trait (Cnaani et al., 2004, Lin et al., 2016, Liu et al., 2014). Novel associations to weight were detected in six linkage groups and may be attributed to either the inclusion of the *O. aureus* genome, or an interaction between the *O. niloticus* and *O. aureus* genomes. However, to date, there have been relatively few studies on weight QTLs in purebred Nile tilapia, blue tilapia, or Nile tilapia x blue tilapia hybrids, and weight should be examined more closely in these species to better understand and unravel the effects of hybridization and of selective breeding on this trait of commercial interest on the genome.

In addition to identifying QTLs for both sex and weight, it would be prudent for future studies to incorporate transcriptomic analyses to explore gene patterns associated with these traits. Dimorphic patterns in gene expression for sex determination would be of particular interest in *O. niloticus* as sex determination in this species is controlled by both genetic and environmental factors (Baroiller et al., 2009, Baroiller et al., 1995, Wessels et al., 2017). Higher temperatures have been shown to induce sex reversal within *O. niloticus* before sexual maturity, i.e. when gonads are still sexually undifferentiated, with genetically female Nile tilapia being phenotypically changed into male tilapia (Baroiller et al., 1995, Lühmann et al. 2012). As males grow faster and larger in Nile tilapia (Alvarado-Ruiz, 2015), this knowledge could be used to increase the production rate of farms using non-hormone induced monosex male cultures and have implications for the entire tilapia industry. Lühmann et al. (2012) found that sex reversal rates were family dependent, suggesting that this may be another trait of interest for the aquaculture industry. As such, it is recommended that transcriptome data be collected from both

male and female gonads at various stages of development during temperature control experiments.

5.6 Understanding Signatures of Selection Between Wild and Domestic Populations

Signatures of selection refers to regions of the genome which are associated, whether directly or in close physical proximity, to genetic variations of traits which have undergone natural or artificial selection (Qanbari et al., 2012, Smith and Haigh, 1974). Understanding these signatures of selection in both wild and domestic populations is advantageous in aquaculture production as it not only allow allows for the understanding of how the genome is affected by domestication, but for improving aquaculture production by identifying wild populations or individuals which already exhibit traits of interest in higher frequencies. These wild populations or individuals can then be used by industry to infuse new genetic material into an already established program while maintaining genetic gain. To ensure that the benefits of these new inclusions is retained in the program, this new germplasm must be actively integrated into the selective breeding program, else their benefits will be short-lived as evidenced in Chapter 2 with the inclusion of a mass spawning line of Nile tilapia into the AS. Additionally, these more advantageous wild populations or individuals can be used to commence a selective breeding program with high performing founders. This would not only result in an increase in production, potentially allowing for the selection line to be disseminated at an earlier date once sufficient genetic gain is achieved, but also improve the maintenance of founder contribution, an important factor to monitor to retain genetic diversity within the program as demonstrated in Chapter 2.

Considering a high number of outlier loci (674) were detected between wild and domestic populations of Nile tilapia, and relatively few (0-11) were detected between wild populations, there is compelling evidence that the majority of these outliers are due to domestication and

animal improvement practices within the AS (Chapter 4). As a *de novo* SNP discovery method was utilized, different marker sets were used for Chapter 3 (linkage mapping and QTL detection) and 4 (outlier detection) whose samples were processed at different dates. However, approximately 30% of identified outliers overlapped with the initial dataset used in Chapters 2 and 3 and could be mapped to at least one of the three (sex average, female, and male) maps created (Chapter 4). Although not all outliers mapped to a QTL region, outliers were identified in five of the six chromosomes exhibiting QTL regions associated with sex and thirteen of the sixteen chromosomes with QTL regions associated with weight in at least one linkage map (Appendix 14). As the AS has undergone selection for weight, detecting outliers located on genomic regions associated with this trait are unsurprising. However, the outliers detected on regions associated with sex within the same species are interesting, and likely reflect the incorporation of the *O. aureus* genome into the AS and its impact on sex determination within the line.

5.7 Determining Wild Population Structuring

The last genetic survey of wild *O. niloticus* populations in Egypt occurred in 2005. It, along with other previous studies, reported greater population structure amongst Nile Delta populations than the present study (Chapter 4; Hassanien and Gilbey, 2005, Hassanien et al., 2004). It is uncertain whether these differences in population structuring are due to differences in molecular technologies utilized by the studies (microsatellites and RAPD; Hassanien and Gilbey, 2005, Hassanien et al., 2004), the dramatic rise in aquaculture in Egypt (Soliman and Yacout, 2016), or a combination of the two. As aquaculture production will only intensify to meet growing demand (FAO, 2018, FAO, 2019a), it is essential to know the current status of wild Nile tilapia populations in order to monitor any future changes and reduce impacts on wild populations.

The population genetic structuring reported within this study describes isolation-by-distance between the Nile Delta and upstream populations of Nile tilapia, with few outliers detected in pairwise comparisons of populations (Chapter 4). Given that the isolation-by-distance also reflects a change from brackish to freshwater environmental conditions, it brings to question whether the observed changes in allele frequencies are due to genetic drift caused by physical distance, or if environmental conditions played a role. Previous GxE experiments for the AS were only conducted in freshwater conditions, for food source and stocking densities (Khaw et al., 2009), not saline conditions. Freshwater populations (Kanater, Aswan, and Asyut) showed higher levels of heterozygosity than in the five Nile Delta populations, possibly suggesting that brackish water populations have some differences in allele frequencies than their riverine counterparts. Conversely, few outliers were detected in pairwise comparisons of Nile Delta and riverine populations (0-11 outliers), suggesting that while allele frequencies may have shifted, alleles have not become fixed in these populations due to selection or that gene flow between the populations is still high enough to counteract selection (Lenormand, 2002).

This is not entirely unexpected as Nile tilapia are renowned for their tolerance to a wide range of environmental conditions (Avella et al., 1993, Balarin and Haller, 1982, Chervinski, 1982, Dominguez et al., 2004, Kamal and Mair, 2005, Philippart and Ruwet, 1982, Randall and Tsui, 2002, Rebouças et al., 2016, Shelton and Popma, 2006). It is possible that these differences in environmental salinity levels may instead be influenced by gene expression as has been observed in other salinity studies of Nile tilapia (Gu et al., 2018, Yamaguchi et al., 2018). While the scope of this study cannot rectify whether allele frequencies, gene flow, gene expression, environmental conditions or a combination of the aforementioned factors can account for the few outliers among wild populations of Nile tilapia, it does highlight that the genetic diversity

required to adapt to various environmental conditions is likely found throughout all wild populations in Egypt. Additional research to understand population differences in salinity tolerance for Nile tilapia and any potential trade-offs with growth rate are highly recommended as sea and estuarine farming of Nile tilapia would expand the production capability and industry potential for this species.

5.8 Identifying the Potential Effects of Dissemination

The goal of most selective breeding lines is dissemination. Prior to dissemination it is advantageous to understand how performance of a line can vary in different locations due to GxE interactions (Callam et al., 2016, Nelson et al., 2018, Proestou et al., 2016) and the impacts that domestic escapees may have on wild populations to maintain the genetic diversity of wild populations.

To understand these interactions running GxE experiments are useful, but not always feasible. This study used wild population structure and signatures of selection through outlier analysis for populations within the dissemination region of interest as a proxy to determine if environments differed within the dissemination area and could have an impact on AS performance. Previous studies of GxE for the AS showed that the lines performed similarly in different environments, indicating that differing environments is not a significant factor in AS performance, at least among freshwater environments (Khaw et al., 2009). Given that Nile tilapia exhibit a wide range of environmental tolerances (Avella et al., 1993, Balarin and Haller, 1982, Chervinski, 1982, Dominguez et al., 2004, Kamal and Mair, 2005, Philippart and Ruwet, 1982, Randall and Tsui, 2002, Rebouças et al., 2016, Shelton and Popma, 2006) and that few outliers (0-11) were detected amongst wild populations, it is suggested that the AS will perform similarly amongst locations once disseminated. However, as some isolation-by-distance and some outliers (0-11)

were detected among brackish Nile Delta populations and freshwater riverine populations it is recommended that targeted GxE studies testing the effect of salinity levels on AS performance and grow-out be conducted in order to understand the potential of expanding production to brackish water.

The AS was shown to have comparable, and in some cases higher, levels of genetic diversity to wild populations of *O. niloticus* (Chapter 4) and low levels of inbreeding (Chapter 2). As all founders were sourced from Egyptian Nile tilapia populations, it is assumed that most of the *O. niloticus* genetic material (90% of available genetic diversity; Grobler, 2017) found within the AS can also be found in wild populations of *O. niloticus* in Egypt, albeit in different frequencies. This is also true for *O. aureus*, which is thought to make up 10% of the AS lines genetic material, (Grobler, 2017) since the Abbassa blue tilapia breeding program was also derived from wild Egyptian stock (Benzie, 2019, pers. comms.). Even though both lines of tilapia have originated from the surrounding wild populations, the effects of hybridized domestic tilapia may have if released within wild populations have not been well explored. However, detrimental effects are not expected within the wild populations as tilapia species are well known for hybridizing in wild environments (D'Amato et al., 2007, Deines et al., 2014, Lovshin, 1982, Meier et al., 2019). While the AS showed the greatest number of polymorphic loci, wild populations all exhibited different subsets of polymorphic loci per sampling location, indicating that hybridization with *O. aureus* may have also occurred in the wild (Chapter 4). Therefore, the introduction of a hybrid escapee is not novel for their species.

The genetic distinction between the AS and wild populations not only allows the detection of putative escapees, but provides the means for long term monitoring of wild populations and the effect of aquaculture escapees can have on them over time. A single putative AS escapee was

detected in this study (Chapter 4), indicating that either escapees are minimal in current management procedures or they are not as fit as their wild counterparts. Escapees, particularly from selectively bred populations, have been shown to have lower fitness in the wild (Yang et al., 2019), as demonstrated by: Atlantic salmon, *Salmo salar* (Glover et al., 2013, McGinnity et al., 2003); European seabass, *Dicentrarchus labrax* (Toledo-Guedes et al., 2014); and Turbot, *Scophthalmus maximus* (Prado et al., 2018). A potential F1 (AS parent x wild parent) escapee was also identified in the wild, suggesting that the AS does have some, albeit seemingly minimal, success in wild populations; however, given that the entirety of genetic diversity within the AS is found within wild populations the AS is not expected to have any significant effect on wild *O. niloticus* in Egypt.

Due to the similar genetic profiles between the AS and wild populations, the AS is a candidate for more widespread dissemination in Egypt. If the line is to be disseminated beyond Egypt, it is recommended that retaining genetic diversity with the line should be a priority for breeders, as this genetic diversity is essential to AS fitness and adaptability to not only new locations, but for changes to future breeding objectives (Notter, 1999). Presently in Egypt, a higher weight at harvest is demanded by the market, but in other locations, or in future generations, other traits such as flesh quality, colour, or disease resistance may also become desirable. If this is the case, there needs to be enough genetic diversity retained within the line for selection of these traits (Notter, 1999).

5.9 Suggestions for Future Aquatic Selective Breeding Programs

Within aquaculture industries, a large amount of information concerning the problems faced within aquatic breeding programs and their solutions are restricted as proprietary data. For example, few on-farm aquatic pedigree error rates, even for well-established salmonoid selective

breeding programs, have been reported despite copious papers dealing with pedigree correction workflows and marker validation (Holman et al., 2017, Liu et al., 2016, Liu et al., 2017, Sellars et al., 2014, Vandeputte et al., 2011). This study only found reference to pedigree error rates in a single conference presentation on a selective breeding program for *Litopenaeus vannamei* available online (Jerry et al., 2017). This lack of transparency has resulted in many programs spending valuable time and resources resolving an issue that may have already been addressed by an external breeding program. This segregation of knowledge is a greater issue in aquaculture than in terrestrial systems, as the logistics of aquatic animals (higher fecundity and smaller in size) allows for the maintenance and success of a program on a single farm vs. the more collaborative efforts observed in terrestrial programs, such as cattle (Bindon, 2001) and sheep (van der Werf et al., 2010), where larger animals require more land. This often results in multiple farms sharing and managing resources, like broodstock.

Given the challenges faced by the Abbassa Strain of Nile tilapia and the findings of this study, it is suggested that the aquaculture industry take the following measures, particularly when establishing a selective breeding program, regardless of species, in order to optimize genetic gain whilst retaining genetic diversity (Table 5.1):

Table 5.10 Suggested measures to be enacted by the aquaculture industry and the outcomes that each measure will have for the program.

Measure	Outcome
1. Collect tissue samples from all broodstock in your program, particularly founders	• Molecular record of initial and available genetic diversity within the line • Can be used to retroactively correct or address problems that arise • Can be used to pursue GS or MAS to improve program productivity
2. Use molecular data to check and correct genealogical records,	• Improved program management • Potential for greater genetic gain as accurate pedigrees yield more improve EBV estimates
3. Optimize founder contribution as a selection criterion,	• Increase retention of genetic diversity • Improves adaptability of program objectives • Reduces the rate of inbreeding accumulation
4. Invest in creating, using, or adapting a SNP Array or similar technology,	• Data directly comparable across generations • Can be used to pursue GS or MAS to improve program productivity
5. Investigate the genetic architecture of traits of interest; and	• Used to pursue GS or MAS to improve program productivity • Used to identify high performing candidate populations or individuals for future breeding programs or to infuse new genetic material into an established line
6. Conduct an audit of wild populations	• Assess dissemination potential • Monitor dissemination effects • Identify high performing candidate populations or individuals for future breeding programs or to infuse new genetic material into an established line
7. Establish a co-operative partnerships	• Share resources- such as SNP Arrays developed for a species • Reduce overall cost of acquiring resources • Improved knowledge of managing and troubleshooting avenues • Overall improved advancement in aquaculture productivity

By following these measures, aquatic selective breeding programs can help ensure that genetic diversity is maintained within their program while maximizing their genetic gain by addressing errors in genealogical records and monitoring founder contribution earlier on in the program. These suggested measures also help establish long-term program goals that can be implemented at a later date. In particular, using a SNP array allows for the same markers to be used throughout the duration of the program, allowing for direct comparisons between generations, the creation of genetic tools that can be built upon, and the pursuit of advanced genetic breeding programs. Additionally, understanding the population structure and genetic diversity within wild populations can help aquaculture breeding programs identify potential risks of dissemination to both the wild population and domestic population performance. It can also allow for the identification of populations or individuals with traits of interest that can be used to infuse a program with new genetic material while minimizing loss of genetic gain or to select broodstock for future breeding programs. It is also recommended that co-operative partnerships, similar to those found in terrestrial systems, or a nucleus breeding facility for implementation and central management of lines be established to share these resources and allow for the advancement of global aquaculture as a whole.

5.10 Conclusions

As a body of work, this thesis has found that both pedigree errors and the incorporation of blue tilapia genetic material into the AS of Nile tilapia have likely contributed to the modest genetic gain observed with the program. Despite this, nominal levels of inbreeding in addition to genetic diversity indices comparable to wild populations indicate that the AS breeding program is recoverable and can be used for further selective breeding. Since the discovery of high pedigree errors and the assessment of the genetic status of the AS, WorldFish is currently addressing the

management issue by bringing in more experienced management, increasing staff training, and employing more stringent practices when collecting genealogical information on farm. To enhance future selective breeding efforts within the line, three novel linkage maps based on sex average, female, and male lines were constructed using a novel population-based methodology, whose use has now been validated for future studies. QTLs associated with both sex and weight were identified within the study; however, given small sample sizes and evidence of hybridization, further studies are required to validate these and determine their relevance to the AS. Despite this, both sex and weight appear to be polygenic in nature and it is recommended that genomic selection be pursued over marker assisted selection for the AS. Wild populations displayed some structuring due to isolation-by-distance; however, few outliers were detected amongst wild populations, suggesting that the AS will perform similarly throughout Egypt once disseminated. Genetic diversity and founder contribution within the line should be maintained not only for the health of the program, but to allow for dissemination of the strain to other regions and for breeding objectives to change with market demand.

REFERENCES

AGHA, S., MEKKAWY, W., IBANEZ-ESCRICHE, N., LIND, C., KUMAR, J., MANDAL, A., BENZIE, J. A. & DOESCHL-WILSON, A. 2018. Breeding for robustness: investigating the genotype-by-environment interaction and micro-environmental sensitivity of Genetically Improved Farmed Tilapia (*Oreochromis niloticus*). *Animal genetics,* 49**,** 421-427.

AGUILAR, I. 2014. SeekParentF90. [Online]. Available: http://nce.ads.uga.edu/wiki/doku.php?id=readme.seekparentf90. [Accessed 11/9/2018]

ALEXANDER, D. H., NOVEMBRE, J. & LANGE, K. 2009. Fast model-based estimation of ancestry in unrelated individuals. *Genome research,* 19**,** 1655-1664.

ALVARADO-RUIZ, C. 2015. Comparison of the growth of males and females of tilapia *Oreochromis niloticus* cultured in cages. *Uniciencia,* 29(1), 1-15.

ANDUEZA-NOH, R. H., MARTÍNEZ-CASTILLO, J. & CHACÓN-SÁNCHEZ, M. I. 2015. Domestication of small-seeded lima bean (*Phaseolus lunatus* L.) landraces in Mesoamerica: evidence from microsatellite markers. *Genetica,* 143**,** 657-669.

ANSAH, Y. B., FRIMPONG, E. A. & HALLERMAN, E. M. 2014. Genetically-improved Tilapia strains in Africa: Potential benefits and negative impacts. *Sustainability,* 6**,** 3697-3721.

ARGUE, B. J., ARCE, S. M., LOTZ, J. M. & MOSS, S. M. 2002. Selective breeding of Pacific white shrimp (*Litopenaeus vannamei*) for growth and resistance to Taura Syndrome Virus. *Aquaculture,* 204**,** 447-460.

ARRUDA, M., LIPKA, A. E., BROWN, P. J., KRILL, A., THURBER, C., BROWN-GUEDIRA, G., DONG, Y., FORESMAN, B. & KOLB, F. L. 2016. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Molecular Breeding,* 36**,** 84.

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**,** 243-254.

AZAZA, M., DHRAIEF, M. & KRAIEM, M. 2008. Effects of water temperature on growth and sex ratio of juvenile Nile tilapia *Oreochromis niloticus* (Linnaeus) reared in geothermal waters in southern Tunisia. *Journal of thermal Biology,* 33**,** 98-105.

BAKER, M. 2012. De novo genome assembly: what every biologist should know. Nature Publishing Group.

BALAH, M. I. 2012. North Delta Lakes, Egypt. *In:* BENGTSSON, L., HERSCHY, R. W. & FAIRBRIDGE, R. W. (eds.) *Encyclopedia of Lakes and Reservoirs.* Dordrecht, Netherlands, Springer.

BALARIN, J. & HALLER, R. 1982. The intensive culture of tilapia in tanks, raceways and cages. *Recent advances in aquaculture,* 1**,** 265-355.

BALDING, D. J., BISHOP, M. J. & CANNINGS, C. 2007. *Handbook of statistical genetics,* Chichester, John Wiley.

BANOS, G., WIGGANS, G. & POWELL, R. 2001. Impact of paternity errors in cow identification on genetic evaluations and international comparisons. *Journal of dairy science,* 84**,** 2523-2529.

BAROILLER, J.-F., D'COTTA, H., BEZAULT, E., WESSELS, S. & HOERSTGEN-SCHWARK, G. 2009. Tilapia sex determination: where temperature and genetics meet. *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology,* 153**,** 30-38.

BAROILLER, J. F., CHOURROUT, D., FOSTIER, A. & JALABERT, B. 1995. Temperature and sex chromosomes govern sex ratios of the mouthbrooding cichlid fish *Oreochromis niloticus*. *Journal of experimental zoology,* 273**,** 216-223.

BENTSEN, H. B. & OLESEN, I. 2002. Designing aquaculture mass selection programs to avoid high inbreeding rates. *Aquaculture,* 204**,** 349-359.

BENTSEN, H. B. & THODESEN, J. 2005. Genetic interactions between farmed and wild fish, with examples from the Atlantic salmon case in Norway. *Selection and breeding programs in aquaculture.* Springer.

BENZIE, J. October 2019 2019. *RE: Program Leader, Sustainable Aquaculture, Research Lead, Kenya at the WorldFish Center.*

BERGERO, R. & CHARLESWORTH, D. 2009. The evolution of restricted recombination in sex chromosomes. *Trends in Ecology & Evolution,* 24**,** 94-102.

BEZAULT, E., ROGNON, X., CLOTA, F., GHARBI, K., BAROILLER, J.-F. & CHEVASSUS, B. 2012. Analysis of the meiotic segregation in intergeneric hybrids of tilapias. *International journal of evolutionary biology,* 2012.

BINDON, B. 2001. Genesis of the Cooperative Research Centre for the Cattle and Beef Industry: integration of resources for beef quality research (1993-2000). *Australian Journal of Experimental Agriculture,* 41**,** 843-853.

BIRO, P. A., ABRAHAMS, M. V., POST, J. R. & PARKINSON, E. A. 2004. Predators select against high growth rates and risk–taking behaviour in domestic trout populations. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 271**,** 2233-2237.

BOICHARD, D. 2002. PEDIG: a fortran package for pedigree analysis suited for large populations. Proceedings of the 7th world congress on genetics applied to livestock production. Montpellier, 525-528.

BOUDRY, P., COLLET, B., CORNETTE, F., HERVOUET, V. & BONHOMME, F. 2002. High variance in reproductive success of the Pacific oyster (*Crassostrea gigas*, Thunberg) revealed by microsatellite-based parentage analysis of multifactorial crosses. *Aquaculture,* 204**,** 283-296.

BOVENHUIS, H. & VAN ARENDONK, J. A. 1991. Estimation of milk protein gene frequencies in crossbred cattle by maximum likelihood. *Journal of dairy science,* 74**,** 2728-2736.

BRADBURY, P. J., ZHANG, Z., KROON, D. E., CASSTEVENS, T. M., RAMDOSS, Y. & BUCKLER, E. S. 2007. TASSEL: software for association mapping of complex traits in diverse samples. *Bioinformatics,* 23**,** 2633-2635.

BRISBANE, J. & GIBSON, J. 1995. Balancing selection response and rate of inbreeding by including genetic relationships in selection decisions. *Theoretical and Applied Genetics,* 91**,** 421-431.

BURKE, J. M., KNAPP, S. J. & RIESEBERG, L. H. 2005. Genetic consequences of selection during the evolution of cultivated sunflower. *Genetics,* 171**,** 1933-1940.

CÁCERES, G., LÓPEZ, M. E., CADIZ, M. I., YOSHIDA, G. M., JEDLICKI, A., PALMA-VÉJARES, R., TRAVISANY, D., DÍAZ-DOMÍNGUEZ, D., MAASS, A., LHORENTE,

J. P., SOTO, J., SALAS, D. & YÁÑEZ, J. M. 2019. Fine mapping using whole-genome sequencing confirms anti-Müllerian hormone as a major gene for sex determination in farmed Nile tilapia (*Oreochromis niloticus* L.). *bioRxiv***,** 573014.

CALLAM, B. R., ALLEN, S. K. & FRANK-LAWALE, A. 2016. Genetic and environmental influence on triploid *Crassostrea virginica* grown in Chesapeake Bay: Growth. *Aquaculture,* 452**,** 97-106.

CAMPOS-RAMOS, R., HARVEY, S. C., MASABANDA, J. S., CARRASCO, L. A., GRIFFIN, D. K., MCANDREW, B., BROMAGE, N. R. & PENMAN, D. J. 2001. Identification of putative sex chromosomes in the blue tilapia, *Oreochromis aureus*, through synaptonemal complex and FISH analysis. *Genetica,* 111**,** 143-153.

CARRASCO, L. A., PENMAN, D. J. & BROMAGE, N. 1999. Evidence for the presence of sex chromosomes in the Nile tilapia (*Oreochromis niloticus*) from synaptonemal complex analysis of XX, XY and YY genotypes. *Aquaculture,* 173**,** 207-218.

CARTER, D., LITI, G., MOSES, A., PARTS, L., JAMES, S., DAVEY, R., ROBERTS, I., BLOMBERG, A., WARRINGER, J. & BURT, A. 2008. Population genomics of domestic and wild yeasts. *Nature Precedings***,** 1-1.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015a. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4**,** s13742-015-0047-8.

CHANG, C. C., CHOW, C. C., TELLIER, L. C., VATTIKUTI, S., PURCELL, S. M. & LEE, J. J. 2015b. Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience,* 4**,** 7.

CHAPUIS, M.-P. & ESTOUP, A. 2006. Microsatellite null alleles and estimation of population differentiation. *Molecular biology and evolution,* 24**,** 621-631.

CHARO-KARISA, H., KOMEN, H., REZK, M. A., PONZONI, R. W., VAN ARENDONK, J. A. & BOVENHUIS, H. 2006. Heritability estimates and response to selection for growth of Nile tilapia (*Oreochromis niloticus*) in low-input earthen ponds. *Aquaculture,* 261**,** 479-486.

CHERVINSKI, J. Environmental physiology of tilapias. 1982. The Biology and Culture of Tilapia. Proceedings of the 7th ICLARM Conference, Manila, Philippines: International Center for Livin, 119-128.

CLUTTON-BROCK, J. 1992. The process of domestication. *Mammal Review,* 22**,** 79-85.

CNAANI, A., ZILBERMAN, N., TINMAN, S., HULATA, G. & RON, M. 2004. Genome-scan analysis for quantitative trait loci in an F 2 tilapia hybrid. *Molecular Genetics and Genomics,* 272**,** 162-172.

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.

COSSU, P., SCARPA, F., SANNA, D., LAI, T., DEDOLA, G. L., CURINI-GALLETTI, M., MURA, L., FOIS, N. & CASU, M. 2019. Influence of genetic drift on patterns of genetic variation: The footprint of aquaculture practices in *Sparus aurata* (Teleostei: Sparidae). *Molecular ecology,* 28**,** 3012-3024.

COURTOIS, B., AUDEBERT, A., DARDOU, A., ROQUES, S., GHNEIM-HERRERA, T., DROC, G., FROUIN, J., ROUAN, L., GOZÉ, E. & KILIAN, A. 2013. Genome-wide association mapping of root traits in a japonica rice panel. *PloS one,* 8**,** e78037.

CRAWFORD, A., TATE, M., MCEWAN, J. & KUMARAMANICKAVEL, G. 1993. How reliable are sheep pedigrees? Proceedings-New Zealand Society of Animal Production. New Zealand Society of Animal Prod Publ, 363-363.
CULLINGHAM, C. I., COOKE, J. E. & COLTMAN, D. W. 2014. Cross-species outlier detection reveals different evolutionary pressures between sister species. *New Phytologist,* 204**,** 215-229.
D'AMATO, M. E., ESTERHUYSE, M. M., VAN DER WAAL, B. C., BRINK, D. & VOLCKAERT, F. A. 2007. Hybridization and phylogeography of the Mozambique tilapia *Oreochromis mossambicus* in southern Africa evidenced by mitochondrial and microsatellite DNA genotyping. *Conservation Genetics,* 8**,** 475-488.
DANZMANN, R. G. 2016. *LINKMFEX: linkage analysis package for outcrossed families with male female exchange of the mapping parent.* [Online]. Available: http://www.uoguelph.ca/~rdanzman/.
DAVID, J. L., HOLTZ, Y. & RANWEZ, V. 2017. The genetic map comparator: a user-friendly application to display and compare genetic maps. *Bioinformatics,* 33**,** 1387-1388.
DAWSON, I. G. & JOHNSON, J. E. 2017. Does size matter? A study of risk perceptions of global population growth. *Risk analysis,* 37**,** 65-81.
DEINES, A., BBOLE, I., KATONGO, C., FEDER, J. & LODGE, D. 2014. Hybridisation between native *Oreochromis* species and introduced Nile tilapia *O. niloticus* in the Kafue River, Zambia. *African Journal of Aquatic Science,* 39**,** 23-34.
DENTON, J. F., LUGO-MARTINEZ, J., TUCKER, A. E., SCHRIDER, D. R., WARREN, W. C. & HAHN, M. W. 2014. Extensive error in the number of genes inferred from draft genome assemblies. *PLoS computational biology,* 10**,** e1003998.
DEY, M. M. & GUPTA, M. V. 2000. Socioeconomics of disseminating genetically improved Nile tilapia in Asia: an introduction.
DO, C., WAPLES, R. S., PEEL, D., MACBETH, G., TILLETT, B. J. & OVENDEN, J. R. 2014. NeEstimator v2: re-implementation of software for the estimation of contemporary effective population size (Ne) from genetic data. *Molecular ecology resources,* 14**,** 209-214.
DOBZHANSKY, T. 1936. Studies on hybrid sterility. II. Localization of sterility factors in Drosophila pseudoobscura hybrids. *Genetics,* 21**,** 113.
DOERKSEN, T. K. & HERBINGER, C. M. 2010. Impact of reconstructed pedigrees on progeny-test breeding values in red spruce. *Tree Genetics & Genomes,* 6**,** 591-600.
DOMINGUEZ, M., TAKEMURA, A., TSUCHIYA, M. & NAKAMURA, S. 2004. Impact of different environmental factors on the circulating immunoglobulin levels in the Nile tilapia, *Oreochromis niloticus*. *Aquaculture,* 241**,** 491-500.
DOYLE, R. M., PEREZ-ENRIQUEZ, R., TAKAGI, M. & TANIGUCHI, N. 2001. Selective recovery of founder genetic diversity in aquacultural broodstocks and captive, endangered fish populations. *Genetica,* 111**,** 291-304.
DU, Q., GONG, C., WANG, Q., ZHOU, D., YANG, H., PAN, W., LI, B. & ZHANG, D. 2016. Genetic architecture of growth traits in Populus revealed by integrated quantitative trait locus (QTL) analysis and association studies. *New Phytologist,* 209**,** 1067-1082.
DUKIĆ, M., BERNER, D., ROESTI, M., HAAG, C. R. & EBERT, D. 2016. A high-density genetic map reveals variation in recombination rate across the genome of *Daphnia magna*. *BMC genetics,* 17**,** 137.

DUPONCHELLE, F. & PANFILI, J. 1998. Variations in age and size at maturity of female Nile tilapia, *Oreochromis niloticus*, populations from man-made lakes of Côte d'Ivoire. *Environmental Biology of Fishes,* 52**,** 453-465.

EARL, D. A. & VONHOLDT, B. M. 2012. STRUCTURE HARVESTER: a website and program for visualizing STRUCTURE output and implementing the Evanno method. *Conservation Genetics Resources,* 4**,** 359-361.

EKNATH, A. & ACOSTA, B. 1998. Genetic improvement of farmed tilapias (GIFT) project: Final report, March 1988 to December 1997.

EKNATH, A., DEY, M., RYE, M., GJERDE, B., ABELLA, T., SEVILLEJA, R., TAYAMEN, M., REYES, R. & BENTSEN, H. 1998. Selective breeding of Nile tilapia for Asia. 6th World Congress on Genetics Applied to Livestock Production, University of New England Armidale, Australia, 89-96.

EKNATH, A. E. & HULATA, G. 2009. Use and exchange of genetic resources of Nile tilapia (Oreochromis niloticus). *Reviews in Aquaculture,* 1**,** 197-213.

EKNATH, A. E., TAYAMEN, M. M., PALADA-DE VERA, M. S., DANTING, J. C., REYES, R. A., DIONISIO, E. E., CAPILI, J. B., BOLIVAR, H. L., ABELLA, T. A. & CIRCA, A. V. 1993. Genetic improvement of farmed tilapias: the growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments. *Genetics in Aquaculture.* Elsevier.

EL-HAWARRY, W. N. 2012. Growth Performance, Proximate Muscle Composition and Dress-Out Percentage of Nile Tilapia (*Oreochromis niloticus*), Blue Tilapia (*Oreochromis aureus*) and their Interspecific Hybrid (*O. aureus* X *O. niloticus*) Cultured in Semi-Intensive Culture System. *World's Vet J,* 2**,** 17-22.

ESHEL, O., SHIRAK, A., WELLER, J., HULATA, G. & RON, M. 2012. Linkage and physical mapping of sex region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3: Genes, Genomes, Genetics,* 2**,** 35-42.

ESHEL, O., SHIRAK, A., WELLER, J., 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**,** 222-224.

EVANNO, G., REGNAUT, S. & GOUDET, J. 2005. Detecting the number of clusters of individuals using the software STRUCTURE: a simulation study. *Molecular ecology,* 14**,** 2611-2620.

EXCOFFIER, L. & LISCHER, H. E. 2010. Arlequin suite ver 3.5: a new series of programs to perform population genetics analyses under Linux and Windows. *Molecular ecology resources,* 10**,** 564-567.

FALCONER, D., MACKAY, T. & FRANKHAM, R. 1996. Selection: I. The response and its prediction. *Introduction to quantitative genetics. Longman, England***,** 184-207.

FALCONER, D. S. 1960. Introduction to quantitative genetics. *Introduction to quantitative genetics.*

FALUSH, D., STEPHENS, M. & PRITCHARD, J. K. 2003. Inference of population structure using multilocus genotype data: linked loci and correlated allele frequencies. *Genetics,* 164**,** 1567-1587.

FALUSH, D., STEPHENS, M. & PRITCHARD, J. K. 2007. Inference of population structure using multilocus genotype data: dominant markers and null alleles. *Molecular ecology notes,* 7**,** 574-578.

FAO 2017. A world overview of species of interest to fisheries.*Oreochromis niloticus. FIGIS Species Fact Sheets* [Online]. ROME: FAO. Available: http://www.fao.org/fishery/culturedspecies/Oreochromis_niloticus/en#tcNA00EA [Accessed 11/9/2017].
FAO 2018. The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals. Rome: Licence: CC BY-NC-SA 3.0 IGO.
FAO 2019a. The State Of The World's Aquatic Genetic Resources for Food and Agriculture. *In:* NATIONS, F. A. A. O. O. T. U. (ed.). Rome.
FAO 2019b. FAO yearbook. Fishery and Aquaculture Statistics 2017. Rome. http://www.fao.org/fishery/statistics/global-aquaculture-production/en.
FAO 2020. The State of World Fisheries and Aquaculture 2020- Sustainability in Action. Rome. https://doi.org/10.4060/ca9229en
FIERST, J. L. 2015. Using linkage maps to correct and scaffold de novo genome assemblies: methods, challenges, and computational tools. *Frontiers in genetics,* 6**,** 220.
FLAQUER, A. & STRAUCH, K. 2012. A comparison of different linkage statistics in small to moderate sized pedigrees with complex diseases. *BMC research notes,* 5**,** 411.
FOLL, M. 2012. BayeScan v2.1 user manual. *Ecology,* 20**,** 1450-1462.
FOLL, M. & GAGGIOTTI, O. 2008. A genome-scan method to identify selected loci appropriate for both dominant and codominant markers: a Bayesian perspective. *Genetics,* 180**,** 977-993.
FRITSCH, P. & RIESEBERG, L. H. 1996. The Use of Random Amplified Polymorphic. *Molecular genetic approaches in conservation,* 56.
FROST, L. A., EVANS, B. S. & JERRY, D. R. 2006. Loss of genetic diversity due to hatchery culture practices in barramundi (*Lates calcarifer*). *Aquaculture,* 261**,** 1056-1064.
GJEDREM, T. & BARANSKI, M. 2010a. Domestication and the Application of Genetic Improvement in Aquaculture. *Selective breeding in aquaculture: an introduction.* Springer Science & Business Media.
GJEDREM, T. & BARANSKI, M. 2010b. *Selective breeding in aquaculture: an introduction*, Springer Science & Business Media.
GJEDREM, T. & BARANSKI, M. 2010c. The Theoretical Basis for Breeding and Selection. *Selective breeding in aquaculture: an introduction.* Springer Science & Business Media.
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**,** 117-129.
GJEDREM, T. & RYE, M. 2018. Selection response in fish and shellfish: a review. *Reviews in Aquaculture,* 10**,** 168-179.
GJERDE, B. 2005. Design of Breeding Programs. *In:* GJEDREM, T. & AKVAFORSK, Å. (eds). *Selection and breeding programs in aquaculture.* Dordrecht, The Netherlands: Springer.
GLOVER, K. A., PERTOLDI, C., BESNIER, F., WENNEVIK, V., KENT, M. & SKAALA, Ø. 2013. Atlantic salmon populations invaded by farmed escapees: quantifying genetic introgression with a Bayesian approach and SNPs. *BMC genetics,* 14**,** 74.
GODDARD, M. & HAYES, B. 2007. Genomic selection. *Journal of Animal breeding and Genetics,* 124**,** 323-330.
GOGARTEN, S. M., BHANGALE, T., CONOMOS, M. P., LAURIE, C. A., MCHUGH, C. P., PAINTER, I., ZHENG, X., CROSSLIN, D. R., LEVINE, D. & LUMLEY, T. 2012.

GWASTools: an R/Bioconductor package for quality control and analysis of genome-wide association studies. *Bioinformatics,* 28**,** 3329-3331.
GONDRO, C., VAN DER WERF, J. & HAYES, B. J. (eds). 2013. *Genome-wide association studies and genomic prediction*. Totowa, NJ, USA: Humana Press.
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**,** 3420-3435.
GOYARD, E., GOARANT, C., ANSQUER, D., BRUN, P., DE DECKER, S., DUFOUR, R., GALINIÉ, C., PEIGNON, J.-M., PHAM, D. & VOUREY, E. 2008. Cross breeding of different domesticated lines as a simple way for genetic improvement in small aquaculture industries: Heterosis and inbreeding effects on growth and survival rates of the Pacific blue shrimp *Penaeus* (*Litopenaeus*) *stylirostris*. *Aquaculture,* 278**,** 43-50.
GRATACAP, R. L., WARGELIUS, A., EDVARDSEN, R. B. & HOUSTON, R. D. 2019. Potential of Genome Editing to Improve Aquaculture Breeding and Production. *Trends in Genetics,* 35**,** 672-684.
GROBLER, M.T. 2017. *Sex determination in the WorldFish Abbassa Strain of Nile Tilapia (Oreochromis niloticus L.).* Master of Science, University of Stirling.
GU, X. H., JIANG, D. L., HUANG, Y., LI, B. J., CHEN, C. H., LIN, H. R. & XIA, J. H. 2018. Identifying a major QTL associated with salinity tolerance in Nile tilapia using QTL-Seq. *Marine biotechnology,* 20**,** 98-107.
GUPTA, M. & ACOSTA, B. 2004. From drawing board to dining table: the success story of the GIFT project. *NAGA, WorldFish Center Quarterly,* 27**,** 4-14.
GUTIERREZ, A., YÁÑEZ, J. & DAVIDSON, W. 2016. Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Marine genomics,* 26**,** 41-50.
GUTIERREZ, A. P., YÁÑEZ, 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.
GUTIÉRREZ, J. A. & GOYACHE, F. 2005. A note on ENDOG: a computer program for analysing pedigree information. *Journal of Animal Breeding and genetics,* 122**,** 172-176.
GUYON, R., RAKOTOMANGA, M., AZZOUZI, N., COUTANCEAU, J. P., BONILLO, C., D'COTTA, H., PEPEY, E., SOLER, L., RODIER-GOUD, M. & D'HONT, A. 2012. A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC genomics,* 13**,** 222.
HASSANIEN, H. A., ELNADY, M., OBEIDA, A. & ITRIBY, H. 2004. Genetic diversity of Nile tilapia populations revealed by randomly amplified polymorphic DNA (RAPD). *Aquaculture Research,* 35**,** 587-593.
HASSANIEN, H. A. & GILBEY, J. 2005. Genetic diversity and differentiation of Nile tilapia (*Oreochromis niloticus*) revealed by DNA microsatellites. *Aquaculture Research,* 36**,** 1450-1457.
HAUG-BALTZELL, A., STEPHENS, S. A., DAVEY, S., SCHEIDEGGER, C. E. & LYONS, E. 2017. SynMap2 and SynMap3D: web-based whole-genome synteny browsers. *Bioinformatics,* 33**,** 2197-2198.
HAYES, B. 2013. Overview of statistical methods for genome-wide association studies (GWAS). *Genome-wide association studies and genomic prediction.* Springer.

HAYES, B. & GODDARD, M. 2010. Genome-wide association and genomic selection in animal breeding. *Genome,* 53**,** 876-883.
HELLER-USZYNSKA, K., USZYNSKI, G., HUTTNER, E., EVERS, M., CARLIG, J., CAIG, V., AITKEN, K., JACKSON, P., PIPERIDIS, G. & COX, M. 2011. Diversity arrays technology effectively reveals DNA polymorphism in a large and complex genome of sugarcane. *Molecular breeding,* 28**,** 37-55.
HIJMANS, R. J., WILLIAMS, E., VENNES, C. & HIJMANS, M. R. J. 2017. Package 'geosphere'. *Spherical trigonometry,* 1,7.
HILL, W., SALISBURY, B. & WEBB, A. 2008. Parentage identification using single nucleotide polymorphism genotypes: application to product tracing. *Journal of animal science,* 86**,** 2508-2517.
HILL, W. G. 2001. Selective Breeding. *In:* BRENNER, S. & MILLER, J. H. (eds.) *Encyclopedia of Genetics.* New York: Academic Press.
HOLMAN, L. E., GARCIA DE LA SERRANA, D., ONOUFRIOU, A., HILLESTAD, B. & JOHNSTON, I. A. 2017. A workflow used to design low density SNP panels for parentage assignment and traceability in aquaculture species and its validation in Atlantic salmon. *Aquaculture,* 476**,** 59-64.
HOSSAIN, S., PANOZZO, J., PITTOCK, C. & FORD, R. 2011. Quantitative trait loci analysis of seed coat color components for selective breeding in chickpea (*Cicer arietinum* L.). *Canadian journal of plant science,* 91**,** 49-55.
HOUSTON, R. D. 2017. Future directions in breeding for disease resistance in aquaculture species. *Revista Brasileira de Zootecnia,* 46**,** 545-551.
HUBBS, C. L. 1955. Hybridization between Fish Species in Nature. *Systematic Biology,* 4**,** 1-20.
HUBISZ, M. J., FALUSH, D., STEPHENS, M. & PRITCHARD, J. K. 2009. Inferring weak population structure with the assistance of sample group information. *Molecular ecology resources,* 9**,** 1322-1332.
HUISMAN, J. 2017. Pedigree reconstruction from SNP data: parentage assignment, sibship clustering and beyond. *Molecular ecology resources,* 17**,** 1009-1024.
IBRAHIM, N. A., ZAID, M. Y. A., KHAW, H. L., EL-NAGGAR, G. O. & PONZONI, R. W. 2013. Relative performance of two Nile tilapia (*Oreochromis niloticus* Linnaeus) strains in Egypt: The Abbassa selection line and the Kafr El Sheikh commercial strain. *Aquaculture Research,* 44**,** 508-517.
ISRAEL, C. & WELLER, J. 2000. Effect of misidentification on genetic gain and estimation of breeding value in dairy cattle populations. *Journal of Dairy Science,* 83**,** 181-187.
JANSSEN, K., CHAVANNE, H., BERENTSEN, P. & KOMEN, H. 2017. Impact of selective breeding on European aquaculture. *Aquaculture,* 472**,** 8-16.
JERRY, D., RAADSMA, H., KHATKAR, M., VAN DER STEEN, H., PROCHASKA, J., JONES, D. & ZENGER, K. 2017. Development of genomic resources and whole genome prediction in Pacific White-Leg Shrimp (*Litopenaeus vannamei*). *Aquaculture*, 107-107.
JOHANSEN-MORRIS, A. & LATTA, R. G. 2006. Fitness consequences of hybridization between ecotypes of *Avena barbata*: hybrid breakdown, hybrid vigor, and transgressive segregation. *Evolution,* 60**,** 1585-1595.
JOHNSSON, J. I., PETERSSON, E., JÖNSSON, E., BJÖRNSSON, B. T. & JÄRVI, T. 1996. Domestication and growth hormone alter antipredator behaviour and growth patterns in juvenile brown trout, *Salmo trutta*. *Canadian Journal of Fisheries and Aquatic Sciences,* 53**,** 1546-1554.

JOMBART, T. 2008. adegenet: a R package for the multivariate analysis of genetic markers. *Bioinformatics,* 24**,** 1403-1405.
JOMBART, T. & AHMED, I. 2011. adegenet 1.3-1: new tools for the analysis of genome-wide SNP data. *Bioinformatics,* 27**,** 3070-3071.
JONES, A. G., SMALL, C. M., PACZOLT, K. A. & RATTERMAN, N. L. 2010. A practical guide to methods of parentage analysis. *Molecular ecology resources,* 10**,** 6-30.
JONES, A. T., OVENDEN, J. R. & WANG, Y. G. 2016. Improved confidence intervals for the linkage disequilibrium method for estimating effective population size. *Heredity,* 117**,** 217-223.
JONES, O. R. & WANG, J. 2010. COLONY: a program for parentage and sibship inference from multilocus genotype data. *Molecular ecology resources,* 10**,** 551-555.
JOSHI, R., ARNYASI, M., LIEN, S., GJØEN, H. M., ALVAREZ, A. T. & KENT, M. 2018. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (*O. niloticus*). *Frontiers in genetics,* 9**,** 472.
KALINOWSKI, S. T. 2011. The computer program STRUCTURE does not reliably identify the main genetic clusters within species: simulations and implications for human population structure. *Heredity,* 106**,** 625.
KALINOWSKI, S. T., TAPER, M. L. & MARSHALL, T. C. 2007. Revising how the computer program CERVUS accommodates genotyping error increases success in paternity assignment. *Molecular ecology,* 16**,** 1099-1106.
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.
KHANAM, T. 2017. Sex determination and genetic management in Nile tilapia using genomic techniques. Doctor of Philosophy, University of Stirling.
KHAW, H. L., BOVENHUIS, H., PONZONI, R. W., REZK, M. A., CHARO-KARISA, H. & KOMEN, H. 2009. Genetic analysis of Nile tilapia (*Oreochromis niloticus*) selection line reared in two input environments. *Aquaculture,* 294**,** 37-42.
KILIAN, A., WENZL, P., HUTTNER, E., CARLING, J., XIA, L., BLOIS, H., CAIG, V., HELLER-USZYNSKA, K., JACCOUD, D. & HOPPER, C. 2012. Diversity arrays technology: a generic genome profiling technology on open platforms. *Data production and analysis in population genomics.* Springer.
KJELDSEN, S. R., RAADSMA, H. W., LEIGH, K. A., TOBEY, J. R., PHALEN, D., KROCKENBERGER, A., ELLIS, W. A., HYNES, E., HIGGINS, D. P. & ZENGER, K. R. 2019. Genomic comparisons reveal biogeographic and anthropogenic impacts in the koala (*Phascolarctos cinereus*): a dietary-specialist species distributed across heterogeneous environments. *Heredity,* 122**,** 525.
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**,** 1225-1232.
KORTE, A. & FARLOW, A. 2013. The advantages and limitations of trait analysis with GWAS: a review. *Plant methods,* 9**,** 29.
KUMAR, N. S. & GURUSUBRAMANIAN, G. 2011. Random amplified polymorphic DNA (RAPD) markers and its applications. *Sci Vis,* 11**,** 116-124.
LADIZINSKY, G. 1985. Founder effect in crop-plant evolution. *Economic Botany,* 39**,** 191-199.

LAWSON, D. J., VAN DORP, L. & FALUSH, D. 2018. A tutorial on how not to over-interpret Structure and Admixture bar plots. *Nature Communications,* 9**,** 3258.
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**,** 237-244.
LEE, B., HULATA, G. & KOCHER, T. 2004. Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity,* 92**,** 543.
LEE, B. Y., PENMAN, D. & KOCHER, T. 2003. Identification of a sex-determining region in Nile tilapia (*Oreochromis niloticus*) using bulked segregant analysis. *Animal genetics,* 34**,** 379-383.
LEE, Y.-S., JEONG, H., TAYE, M., KIM, H. J., KA, S., RYU, Y.-C. & CHO, S. 2015. Genome-wide association study (GWAS) and its application for improving the genomic estimated breeding values (GEBV) of the Berkshire pork quality traits. *Asian-Australasian journal of animal sciences,* 28**,** 1551.
LENORMAND, T. 2002. Gene flow and the limits to natural selection. *Trends in Ecology & Evolution,* 17**,** 183-189.
LI, W. & GODZIK, A. 2006. Cd-hit: a fast program for clustering and comparing large sets of protein or nucleotide sequences. *Bioinformatics,* 22**,** 1658-1659.
LIN, G., CHUA, E., ORBAN, L. & YUE, G. H. 2016. Mapping QTL for sex and growth traits in salt-tolerant tilapia (*Oreochromis* spp. X *O. mossambicus*). *PloS one,* 11**,** e0166723.
LIND, C., KILIAN, A. & BENZIE, J. 2017. Development of diversity arrays technology markers as a tool for rapid genomic assessment in Nile tilapia, *Oreochromis niloticus*. *Animal genetics,* 48**,** 362-364.
LIND, C., PONZONI, R., NGUYEN, N. & KHAW, H. 2012. Selective Breeding in Fish and Conservation of Genetic Resources for Aquaculture. *Reproduction in Domestic Animals,* 47**,** 255-263.
LIND, C. E., EVANS, B. S., TAYLOR, J. J. & JERRY, D. R. 2010. The consequences of differential family survival rates and equalizing maternal contributions on the effective population size (Ne) of cultured silver-lipped pearl oysters, *Pinctada maxima*. *Aquaculture research,* 41**,** 1229-1242.
LINLØKKEN, A. N., HAUGEN, T. O., KENT, M. P. & LIEN, S. 2017. Genetic differences between wild and hatchery-bred brown trout (*Salmo trutta* L.) in single nucleotide polymorphisms linked to selective traits. *Ecology and evolution,* 7**,** 4963-4972.
LITTLE, D. C., NEWTON, R. & BEVERIDGE, M. 2016. Aquaculture: a rapidly growing and significant source of sustainable food? Status, transitions and potential. *Proceedings of the Nutrition Society,* 75**,** 274-286.
LITTRELL, J., TSAIH, S.-W., BAUD, A., RASTAS, P., SOLBERG-WOODS, L. & FLISTER, M. J. 2018. A High-Resolution Genetic Map for the Laboratory Rat. *G3: Genes Genomes, Genetics,* 8**,** 2241-2248.
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. 2014. A genome scan revealed significant associations of growth traits with a major QTL and GHR2 in tilapia. *Scientific reports,* 4**,** 7256.
LIU, S., PALTI, Y., GAO, G. & REXROAD, C. E. 2016. Development and validation of a SNP panel for parentage assignment in rainbow trout. *Aquaculture,* 452**,** 178-182.
LIU, S., ZHOU, Z., LU, J., SUN, F., WANG, S., LIU, H., JIANG, Y., KUCUKTAS, H., KALTENBOECK, L., PEATMAN, E. & LIU, Z. 2011. Generation of genome-scale

gene-associated SNPs in catfish for the construction of a high-density SNP array. *BMC Genomics,* 12**,** 53.
LIU, T., LI, Q., KONG, L. & YU, H. 2017. Comparison of microsatellites and SNPs for pedigree analysis in the Pacific oyster *Crassostrea gigas*. *Aquaculture International,* 25**,** 1507-1519.
LIU, Z. J. 2017. *Bioinformatics in Aquaculture: Principles and Methods*, John Wiley & Sons.
LÓPEZ, M. E., BENESTAN, L., MOORE, J. S., PERRIER, C., GILBEY, J., DI GENOVA, A., MAASS, A., DIAZ, D., LHORENTE, J. P. & CORREA, K. 2019. Comparing genomic signatures of domestication in two Atlantic salmon (*Salmo salar* L.) populations with different geographical origins. *Evolutionary applications,* 12**,** 137-156.
LOUGHNAN, S. R., SMITH-KEUNE, C., JERRY, D. R., BEHEREGARAY, L. B. & ROBINSON, N. A. 2016. Genetic diversity and relatedness estimates for captive barramundi (*Lates calcarifer*, Bloch) broodstock informs efforts to form a base population for selective breeding. *Aquaculture Research,* 47**,** 3570-3584.
LOVSHIN, L. 1982 Tilapia hybridization. International Conference on the Biology and Culture of Tilapias, Bellagio (Italy), 2-5 Sep 1980.
LÜHMANN, L.M., KNORR, C., HŐRSTGEN-SCHWARK, G., WESSELS, S. 2012. First evidence for family-specific QTL for temperature-dependent sex reversal in Nile tilapia (*Oreochromis niloticus*). *Sexual Development*, 6(5), 247-256.
LYONS, E., PEDERSEN, B., KANE, J. & FREELING, M. 2008. The value of nonmodel genomes and an example using SynMap within CoGe to dissect the hexaploidy that predates the rosids. *Tropical Plant Biology,* 1**,** 181-190.
MAIR, G., SCOTT, A., PENMAN, D., BEARDMORE, J. & SKIBINSKI, D. 1991. Sex determination in the genus *Oreochromis*. *Theoretical and Applied Genetics,* 82**,** 144-152.
MAKINO, T., RUBIN, C.-J., CARNEIRO, M., AXELSSON, E., ANDERSSON, L. & WEBSTER, M. T. 2018. Elevated Proportions of Deleterious Genetic Variation in Domestic Animals and Plants. *Genome Biology and Evolution,* 10**,** 276-290.
MANK, J. & AVISE, J. 2009. Evolutionary diversity and turn-over of sex determination in teleost fishes. *Sexual Development,* 3**,** 60-67.
MANK, J. E. 2008. Sex chromosomes and the evolution of sexual dimorphism: lessons from the genome. *The American Naturalist,* 173**,** 141-150.
MARRANO, A., MICHELETTI, D., LORENZI, S., NEALE, D. & GRANDO, M. S. 2018. Genomic signatures of different adaptations to environmental stimuli between wild and cultivated *Vitis vinifera* L. *Horticulture research,* 5**,** 34.
MCGINNITY, P., PRODÖHL, P., FERGUSON, A., HYNES, R., MAOILÉIDIGH, N. Ó., BAKER, N., COTTER, D., O'HEA, B., COOKE, D. & ROGAN, G. 2003. Fitness reduction and potential extinction of wild populations of Atlantic salmon, *Salmo salar*, as a result of interactions with escaped farm salmon. *Proceedings of the Royal Society of London. Series B: Biological Sciences,* 270**,** 2443-2450.
MEIER, J. I., MARQUES, D. A., MWAIKO, S., WAGNER, C. E., EXCOFFIER, L. & SEEHAUSEN, O. 2017. Ancient hybridization fuels rapid cichlid fish adaptive radiations. *Nature communications,* 8**,** 1-11.
MEIER, J. I., STELKENS, R. B., JOYCE, D. A., MWAIKO, S., PHIRI, N., SCHLIEWEN, U. K., SELZ, O. M., WAGNER, C. E., KATONGO, C. & SEEHAUSEN, O. 2019. The coincidence of ecological opportunity with hybridization explains rapid adaptive radiation in Lake Mweru cichlid fishes. *Nature Communications,* 10**,** 1-11.

MIA, M. Y., TAGGART, J. B., GILMOUR, A. E., GHEYAS, A. A., DAS, T. K., KOHINOOR, A. H. M., RAHMAN, M. A., SATTAR, M. A., HUSSAIN, M. G., MAZID, M. A., PENMAN, D. J. & MCANDREW, B. J. 2005. Detection of hybridization between Chinese carp species (*Hypophthalmichthys molitrix* and *Aristichthys nobilis*) in hatchery broodstock in Bangladesh, using DNA microsatellite loci. *Aquaculture,* 247**,** 267-273.

MIGNON-GRASTEAU, S., BOISSY, A., BOUIX, J., FAURE, J.-M., FISHER, A. D., HINCH, G. N., JENSEN, P., LE NEINDRE, P., MORMEDE, P. & PRUNET, P. 2005. Genetics of adaptation and domestication in livestock. *Livestock Production Science,* 93**,** 3-14.

MOAWAD, M. B., ABDEL AZIZ, A. O. & MAMTIMIN, B. 2016. Flash floods in the Sahara: a case study for the 28 January 2013 flood in Qena, Egypt. *Geomatics, Natural Hazards and Risk,* 7**,** 215-236.

MOSS, S. M., MOSS, D. R., ARCE, S. M., LIGHTNER, D. V. & LOTZ, J. M. 2012. The role of selective breeding and biosecurity in the prevention of disease in penaeid shrimp aquaculture. *Journal of invertebrate pathology,* 110**,** 247-250.

MSANGI, S., KOBAYASHI, M., BATKA, M., VANNUCCINI, S., DEY, M. & ANDERSON, J. 2013. Fish to 2030: prospects for fisheries and aquaculture. *World Bank Report,* 83177**,** 102.

MUNOZ, P. R., RESENDE, M. F., HUBER, D. A., QUESADA, T., RESENDE, M. D., NEALE, D. B., WEGRZYN, J. L., KIRST, M. & PETER, G. F. 2014. Genomic relationship matrix for correcting pedigree errors in breeding populations: impact on genetic parameters and genomic selection accuracy. *Crop Science,* 54**,** 1115-1123.

NAYFA, M. G., JONES, D. B., LIND, C. E., BENZIE, J. A. H., JERRY, D. R. & ZENGER, K. R. 2020. Pipette and paper: Combining molecular and genealogical methods to assess a Nile tilapia (*Oreochromis niloticus*) breeding program. *Aquaculture,* 523**,** 735171.

NAYFA, M. G. & ZENGER, K. R. 2016. Unravelling the effects of gene flow and selection in highly connected populations of the silver-lip pearl oyster (*Pinctada maxima*). *Marine genomics,* 28**,** 99-106.

NCBI Resources Coordinatiors. 2017. Database resources of the national center for biotechnology information. *Nucleic acids researcher,* 45(D1),12-17.

NELSON, N. D., BERGUSON, W. E., MCMAHON, B. G., CAI, M. & BUCHMAN, D. J. 2018. Growth performance and stability of hybrid poplar clones in simultaneous tests on six sites. *Biomass and Bioenergy,* 118**,** 115-125.

NEUDITSCHKO, M., KHATKAR, M. S. & RAADSMA, H. W. 2012. NetView: a high-definition network-visualization approach to detect fine-scale population structures from genome-wide patterns of variation. *PloS one,* 7**,** e48375.

NGUYEN, N. H. 2016. Genetic improvement for important farmed aquaculture species with a reference to carp, tilapia and prawns in Asia: achievements, lessons and challenges. *Fish and Fisheries,* 17**,** 483-506.

NIVELLE, R., GENNOTTE, V., KALALA, E. J. K., NGOC, N. B., MULLER, M., MELARD, C. & ROUGEOT, C. 2019. Temperature preference of Nile tilapia (Oreochromis niloticus) juveniles induces spontaneous sex reversal. *PloS one,* 14.

NOTTER, D. R. 1999. The importance of genetic diversity in livestock populations of the future. *Journal of animal science,* 77**,** 61-69.

NWOGWUGWU, C. P., KIM, Y., CHUNG, Y. J., JANG, S. B., ROH, S. H., KIM, S., LEE, J. H., CHOI, T. J. & LEE, S. H. 2019. Effect of errors in pedigree on the accuracy of

estimated breeding value for carcass traits in Korean Hanwoo cattle. *Asian-Australasian Journal of Animal Sciences*.
OLLIVIER, L. 2002. *Eléments de génétique quantitative: 2e édition revue et augmentée*, Editions Quae.
OUEDRAOGO, C., CANONNE, M., D'COTTA, H., BAROILLER, J.-F. & BARAS, E. 2014. Minimal body size for tagging fish with electronic microchips as studied in the Nile Tilapia. *North American Journal of Aquaculture,* 76**,** 275-280.
OUELLETTE, L. A., REID, R. W., BLANCHARD, S. G. & BROUWER, C. R. 2017. LinkageMapView—rendering high-resolution linkage and QTL maps. *Bioinformatics,* 34**,** 306-307.
PADHI, B. & MANDAL, R. 1997. Inadvertent hybridization in a carp hatchery as detected by nuclear DNA RFLP. *Journal of Fish Biology,* 50**,** 906-909.
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., KHAN, M. G., TAGGART, J. B., GHARBI, K., MCANDREW, B. J. & PENMAN, D. J. 2015. A novel sex-determining QTL in Nile tilapia (*Oreochromis niloticus*). *BMC Genomics,* 16**,** 171.
PANTE, M. J. R., GJERDE, B. & MCMILLAN, I. 2001. Inbreeding levels in selected populations of rainbow trout, *Oncorhynchus mykiss*. *Aquaculture,* 192**,** 213-224.
PAQUETTE, S. R. & PAQUETTE, M. S. R. 2011. Package 'PopGenKit'.
PAYSEUR, B. A. & RIESEBERG, L. H. 2016. A genomic perspective on hybridization and speciation. *Molecular Ecology,* 25**,** 2337-2360.
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.
PHILIPPART, J.C. & RUWET, J.C. 1982. Ecology and distribution of tilapias. *The biology and culture of tilapias,* 7**,** 15-60.
PLATT, S. & HAUSER, W. J. 1978. Optimum temperature for feeding and growth of Tilapia zillii. *The Progressive Fish-Culturist,* 40**,** 105-107.
PONZONI, R., NGUYEN, N. H., KHAW, H. L., KAMARUZZAMAN, N., HAMZAH, A., BAKAR, K. A. & YEE, H. 2008. Genetic improvement of Nile tilapia (*Oreochromis niloticus*)—Present and future. *International Symposium on Tilapia in Aquaculture,* 33-52.
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.
POWELL, J. E., VISSCHER, P. M. & GODDARD, M. E. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. *Nature Reviews Genetics,* 11**,** 800.
PRADO, F., VERA, M., HERMIDA, M., BLANCO, A., BOUZA, C., MAES, G., VOLCKAERT, F., MARTÍNEZ, P. & CONSORTIUM, A. 2018. Tracing the genetic impact of farmed turbot *Scophthalmus maximus* on wild populations. *Aquaculture Environment Interactions,* 10**,** 447-463.

PRICE, E. O. 1984. Behavioral aspects of animal domestication. *The quarterly review of biology,* 59**,** 1-32.
PRITCHARD, J. K., STEPHENS, M. & DONNELLY, P. 2000. Inference of population structure using multilocus genotype data. *Genetics,* 155**,** 945-959.
PROESTOU, D. A., VINYARD, B. T., CORBETT, R. J., PIESZ, J., ALLEN, S. K., SMALL, J. M., LI, C., LIU, M., DEBROSSE, G., GUO, X., RAWSON, P. & GÓMEZ-CHIARRI, M. 2016. Performance of selectively-bred lines of eastern oyster, *Crassostrea virginica*, across eastern US estuaries. *Aquaculture,* 464**,** 17-27.
PURCELL, S. & CHANG, C. 2017. PLINK 1.90 beta. Available: https://www.cog-genomics.org/plink2. [Accessed 13/3/2018].
PURCELL, S., NEALE, B., TODD-BROWN, K., THOMAS, L., FERREIRA, M. A., BENDER, D., MALLER, J., SKLAR, P., DE BAKKER, P. I. & DALY, M. J. 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. *The American journal of human genetics,* 81**,** 559-575.
PUTMAN, A. I. & CARBONE, I. 2014. Challenges in analysis and interpretation of microsatellite data for population genetic studies. *Ecology and evolution,* 4**,** 4399-4428.
QANBARI, S., STROM, T. M., HABERER, G., WEIGEND, S., GHEYAS, A. A., TURNER, F., BURT, D. W., PREISINGER, R., GIANOLA, D. & SIMIANER, H. 2012. A high resolution genome-wide scan for significant selective sweeps: an application to pooled sequence data in laying chickens. *PloS one,* 7.
R CORE TEAM 2018. R: A Language and Environment for Statistical Computing. *In:* COMPUTING, R. F. F. S. (ed.). Vienna, Austria.
RAHMAN, M. A., ARSHAD, A., MARIMUTHU, K., ARA, R. & AMIN, S. 2013. Inter-specific hybridization and its potential for aquaculture of fin fishes. *Asian J. Anim. Vet. Adv,* 8**,** 139-153.
RAMAN, H., RAMAN, R., KILIAN, A., DETERING, F., CARLING, J., COOMBES, N., DIFFEY, S., KADKOL, G., EDWARDS, D. & MCCULLY, M. 2014. Genome-wide delineation of natural variation for pod shatter resistance in *Brassica napus*. *PLoS One,* 9**,** e101673.
RANA, K. 1988. Reproductive biology and the hatchery rearing of tilapia eggs and fry. *Recent advances aquaculture,* 343-406. Springer, Dordrecht.
RANDALL, D. J. & TSUI, T. 2002. Ammonia toxicity in fish. *Marine pollution bulletin,* 45**,** 17-23.
REBOUÇAS, V. T., LIMA, F. R. D. S. & CAVALCANTE, D. D. H. 2016. Reassessment of the suitable range of water pH for culture of Nile tilapia *Oreochromis niloticus* L. in eutrophic water. *Acta Scientiarum. Animal Sciences,* 38**,** 361-368.
REZK, M. A., PONZONI, R. W., KHAW, H. L., KAMEL, E., DAWOOD, T. & JOHN, G. 2009. Selective breeding for increased body weight in a synthetic breed of Egyptian Nile tilapia, *Oreochromis niloticus*: response to selection and genetic parameters. *Aquaculture,* 293**,** 187-194.
ROBINSON, N. & HAYES, B. 2008. Modelling the use of gene expression profiles with selective breeding for improved disease resistance in Atlantic salmon (*Salmo salar*). *Aquaculture,* 285**,** 38-46.
ROBLEDO, D., PALAIOKOSTAS, C., BARGELLONI, L., MARTÍNEZ, P. & HOUSTON, R. 2018. Applications of genotyping by sequencing in aquaculture breeding and genetics. *Reviews in aquaculture,* 10**,** 670-682.

RYMAN, N. & LAIKRE, L. 1991. Effects of Supportive Breeding on the Genetically Effective Population Size. *Conservation Biology,* 5**,** 325-329.
SALIN, D. & WILLIOT, P. 1991. Acute toxicity of ammonia to Siberian sturgeon, *Acipenser baeri*.*Williot, Ed. Acipenser Cemagref Publ***,** 153-167.
SANDERS, K., BENNEWITZ, J. & KALM, E. 2006. Wrong and missing sire information affects genetic gain in the Angeln dairy cattle population. *Journal of dairy science,* 89**,** 315-321.
SCANDURA, M., IACOLINA, L. & APOLLONIO, M. 2011. Genetic diversity in the European wild boar *Sus scrofa*: phylogeography, population structure and wild x domestic hybridization. *Mammal review,* 41**,** 125-137.
SCHUMER, M., CUI, R., POWELL, D. L., DRESNER, R., ROSENTHAL, G. G. & ANDOLFATTO, P. 2014. High-resolution mapping reveals hundreds of genetic incompatibilities in hybridizing fish species. *Elife,* 3**,** e02535.
SCHUMER, M., XU, C., POWELL, D. L., DURVASULA, A., SKOV, L., HOLLAND, C., BLAZIER, J. C., SANKARARAMAN, S., ANDOLFATTO, P. & ROSENTHAL, G. G. 2018. Natural selection interacts with recombination to shape the evolution of hybrid genomes. *Science,* 360**,** 656-660.
SCRIBNER, K. T., PAGE, K. S. & BARTRON, M. L. 2000. Hybridization in freshwater fishes: a review of case studies and cytonuclear methods of biological inference. *Reviews in Fish Biology and Fisheries,* 10**,** 293-323.
SELLARS, M. J., DIERENS, L., MCWILLIAM, S., LITTLE, B., MURPHY, B., COMAN, G. J., BARENDSE, W. & HENSHALL, J. 2014. Comparison of microsatellite and SNP DNA markers for pedigree assignment in Black Tiger shrimp, *Penaeus monodon*. *Aquaculture Research,* 45**,** 417-426.
SHELTON, W. L. & POPMA, T. J. 2006. Biology. *In:* WEBSTER, C. D. & LIM, C. (eds.) *Tilapia: biology, culture, and nutrition.* CRC Press.
SHIRAK, A., ZAK, T., DOR, L., BENET-PERLBERG, A., WELLER, J. I., RON, M. & SEROUSSI, E. 2019. Quantitative trait loci on LGs 9 and 14 affect the reproductive interaction between two *Oreochromis* species, *O. niloticus* and *O. aureus*. *Heredity,* 122**,** 341-353.
SILVA-JUNIOR, O. B., FARIA, D. A. & GRATTAPAGLIA, D. 2015. A flexible multi-species genome-wide 60K SNP chip developed from pooled resequencing of 240 Eucalyptus tree genomes across 12 species. *New Phytologist,* 206**,** 1527-1540.
SIMMONS, M., MICKETT, K., KUCUKTAS, H., LI, P., DUNHAM, R. & LIU, Z. 2006. Comparison of domestic and wild channel catfish (*Ictalurus punctatus*) populations provides no evidence for genetic impact. *Aquaculture,* 252**,** 133-146.
SKAARUD, A., WOOLLIAMS, J. A. & GJØEN, H. M. 2014. Optimising resources and management of genetic variation in fish-breeding schemes with multiple traits. *Aquaculture,* 420**,** 133-138.
SMITH, J. M. & HAIGH, J. 1974. The hitch-hiking effect of a favourable gene. *Genetics Research,* 23**,** 23-35.
SOLIMAN, N. F. & YACOUT, D. M. 2016. Aquaculture in Egypt: status, constraints and potentials. *Aquaculture international,* 24**,** 1201-1227.
SONESSON, A. K., WOOLLIAMS, J. A. & MEUWISSEN, T. H. 2012. Genomic selection requires genomic control of inbreeding. *Genetics Selection Evolution,* 44**,** 27.

STAM, P. 1993. Construction of integrated genetic linkage maps by means of a new computer package: Join Map. *The Plant Journal,* 3**,** 739-744.
STEINIG, E. J., NEUDITSCHKO, M., KHATKAR, M. S., RAADSMA, H. W. & ZENGER, K. R. 2016. netview p: a network visualization tool to unravel complex population structure using genome-wide SNPs. *Molecular Ecology Resources,* 16**,** 216-227.
STELKENS, R. B., SCHMID, C. & SEEHAUSEN, O. 2015. Hybrid breakdown in cichlid fish. *PloS one,* 10.
STOFFEL, M. A., ESSER, M., KARDOS, M., HUMBLE, E., NICHOLS, H., DAVID, P. & HOFFMAN, J. I. 2016. inbreedR: an R package for the analysis of inbreeding based on genetic markers. *Methods in Ecology and Evolution,* 7**,** 1331-1339.
STRONEN, A. V., SALMELA, E., BALDURSDOTTIR, B. K., BERG, P., ESPELIEN, I. S., JÄRVI, K., JENSEN, H., KRISTENSEN, T. N., MELIS, C. & MANENTI, T. 2017. Genetic rescue of an endangered domestic animal through outcrossing with closely related breeds: A case study of the Norwegian Lundehund. *PloS one,* 12.
TOLEDO-GUEDES, K., SANCHEZ-JEREZ, P., BENJUMEA, M. E. & BRITO, A. 2014. Farming-up coastal fish assemblages through a massive aquaculture escape event. *Marine environmental research,* 98**,** 86-95.
TSAI, H.-Y., HAMILTON, A., TINCH, A. E., GUY, D. R., GHARBI, K., STEAR, M. J., MATIKA, O., BISHOP, S. C. & HOUSTON, R. D. 2015. Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics,* 16**,** 969.
TURNER, S. D. 2014. qqman: an R package for visualizing GWAS results using Q-Q and manhattan plots. *bioRxiv***,** 005165.
VAN DER WERF, J. H. J., KINGHORN, B. P. & BANKS, R. G. 2010. Design and role of an information nucleus in sheep breeding programs. *Animal Production Science,* 50**,** 998-1003.
VAN OOIJEN, J. 2009. MapQTL 6, Software for the mapping of quantitative trait loci in experimental populations of diploid species. *Kyazma BV, Wageningen, The Netherlands*.
VAN OOIJEN, J. 2018. JoinMap 5: Software for the calculation of genetic linkage maps in experimental populations of diploid species. *Kyazma BV, Wageningen, The Netherlands*.
VANDEPUTTE, M., ROSSIGNOL, M.-N. & PINCENT, C. 2011. From theory to practice: empirical evaluation of the assignment power of marker sets for pedigree analysis in fish breeding. *Aquaculture,* 314**,** 80-86.
VISSCHER, P., WOOLLIAMS, J., SMITH, D. & WILLIAMS, J. 2002. Estimation of pedigree errors in the UK dairy population using microsatellite markers and the impact on selection. *Journal of dairy science,* 85**,** 2368-2375.
VISSCHER, P. M., HEMANI, G., VINKHUYZEN, A. A., CHEN, G.-B., LEE, S. H., WRAY, N. R., GODDARD, M. E. & YANG, J. 2014. Statistical power to detect genetic (co) variance of complex traits using SNP data in unrelated samples. *PLoS genetics,* 10**,** e1004269.
VON MARK, V. C., KILIAN, A. & DIERIG, D. A. 2013. Development of DArT marker platforms and genetic diversity assessment of the US collection of the new oilseed crop *Lesquerella* and related species. *PLoS one,* 8**,** e64062.
WAN, S., LI, Q., LIU, T., YU, H. & KONG, L. 2017. Heritability estimates for shell color-related traits in the golden shell strain of Pacific oyster (*Crassostrea gigas*) using a molecular pedigree. *Aquaculture,* 476**,** 65-71.

WANG, J., LIU, Y., JIANG, S., LI, W., GUI, L., ZHOU, T., ZHAI, W., LIN, Z., LU, J. & CHEN, L. 2019. Transcriptomic and epigenomic alterations of Nile tilapia gonads sexually reversed by high temperature. *Aquaculture,* 508**,** 167-177.

WAPLES, R. S. 2014. Testing for Hardy–Weinberg Proportions: Have We Lost the Plot? *Journal of Heredity,* 106**,** 1-19.

WAPLES, R. S. & GAGGIOTTI, O. 2006. INVITED REVIEW: What is a population? An empirical evaluation of some genetic methods for identifying the number of gene pools and their degree of connectivity. *Molecular ecology,* 15**,** 1419-1439.

WESSELS, S., KRAUSE, I., FLOREN, C., SCHÜTZ, E., BECK, J. & KNORR, C. 2017. ddRADseq reveals determinants for temperature-dependent sex reversal in Nile tilapia on LG23. *BMC genomics,* 18**,** 531.

WESSELS, S., SHARIFI, R. A., LUEHMANN, L. M., RUEANGSRI, S., KRAUSE, I., PACH, S., HOERSTGEN-SCHWARK, G. & KNORR, C. 2014. Allelic variant in the anti-müllerian hormone gene leads to autosomal and temperature-dependent sex reversal in a selected Nile tilapia line. *PloS one,* 9.

WILLIS, P. M. 2013. Why do animals hybridize? *Acta Ethologica,* 16**,** 127-134.

WOOLLIAMS, J. 1994. Effective sizes of livestock populations to prevent a decline in fitness. *Theoretical and Applied Genetics,* 89**,** 1019-1026.

WORLDFISH. 2016. *The GIFT that keeps giving: Genetically Improved Farmed Tilapia (GIFT) has been developed for nearly 30 years to have fast growth, benefiting millions across the world.* [Online]. Available: https://www.worldfishcenter.org/pages/gift/ [Accessed 15/10/2019].

YAMAGUCHI, Y., BREVES, J. P., HAWS, M. C., LERNER, D. T., GRAU, E. G. & SEALE, A. P. 2018. Acute salinity tolerance and the control of two prolactins and their receptors in the Nile tilapia (*Oreochromis niloticus*) and Mozambique tilapia (*O. mossambicus*): a comparative study. *General and comparative endocrinology,* 257**,** 168-176.

YANG, J., LEE, S. H., GODDARD, M. E. & VISSCHER, P. M. 2011. GCTA: a tool for genome-wide complex trait analysis. *The American Journal of Human Genetics,* 88**,** 76-82.

YANG, J., ZAITLEN, N. A., GODDARD, M. E., VISSCHER, P. M. & PRICE, A. L. 2014. Advantages and pitfalls in the application of mixed-model association methods. *Nature genetics,* 46**,** 100.

YANG, L., WAPLES, R. S. & BASKETT, M. L. 2019. Life history and temporal variability of escape events interactively determine the fitness consequences of aquaculture escapees on wild populations. *Theoretical population biology*. 129, 93-102.

ZAMANI, W., GHASEMPOURI, S. M., REZAEI, H. R., NADERI, S., HESARI, A. R. E. & OUHROUCH, A. 2018. Comparing polymorphism of 86 candidate genes putatively involved in domestication of sheep, between wild and domestic Iranian sheep. *Meta Gene,* 17**,** 223-231.

ZENGER, K., KHATKAR, M., JERRY, D. & RAADSMA, H. 2017. The next wave in selective breeding: implementing genomic selection in aquaculture. *Proc. Assoc. Advmt. Anim. Breed. Genet*, 22, 105-112.

ZENGER, K. R., KHATKAR, M. S., JONES, D. B., KHALILISAMANI, N., JERRY, D. R. & RAADSMA, H. W. 2019. Genomic selection in aquaculture: application, limitations and opportunities with special reference to marine shrimp and pearl oysters. *Frontiers in genetics,* 9**,** 693.

ZHENG, H., ZHANG, T., SUN, Z., LIU, W. & LIU, H. 2013. Inheritance of shell colours in the noble scallop *Chlamys nobilis* (Bivalve: Pectinidae). *Aquaculture Research,* 44**,** 1229-1235.

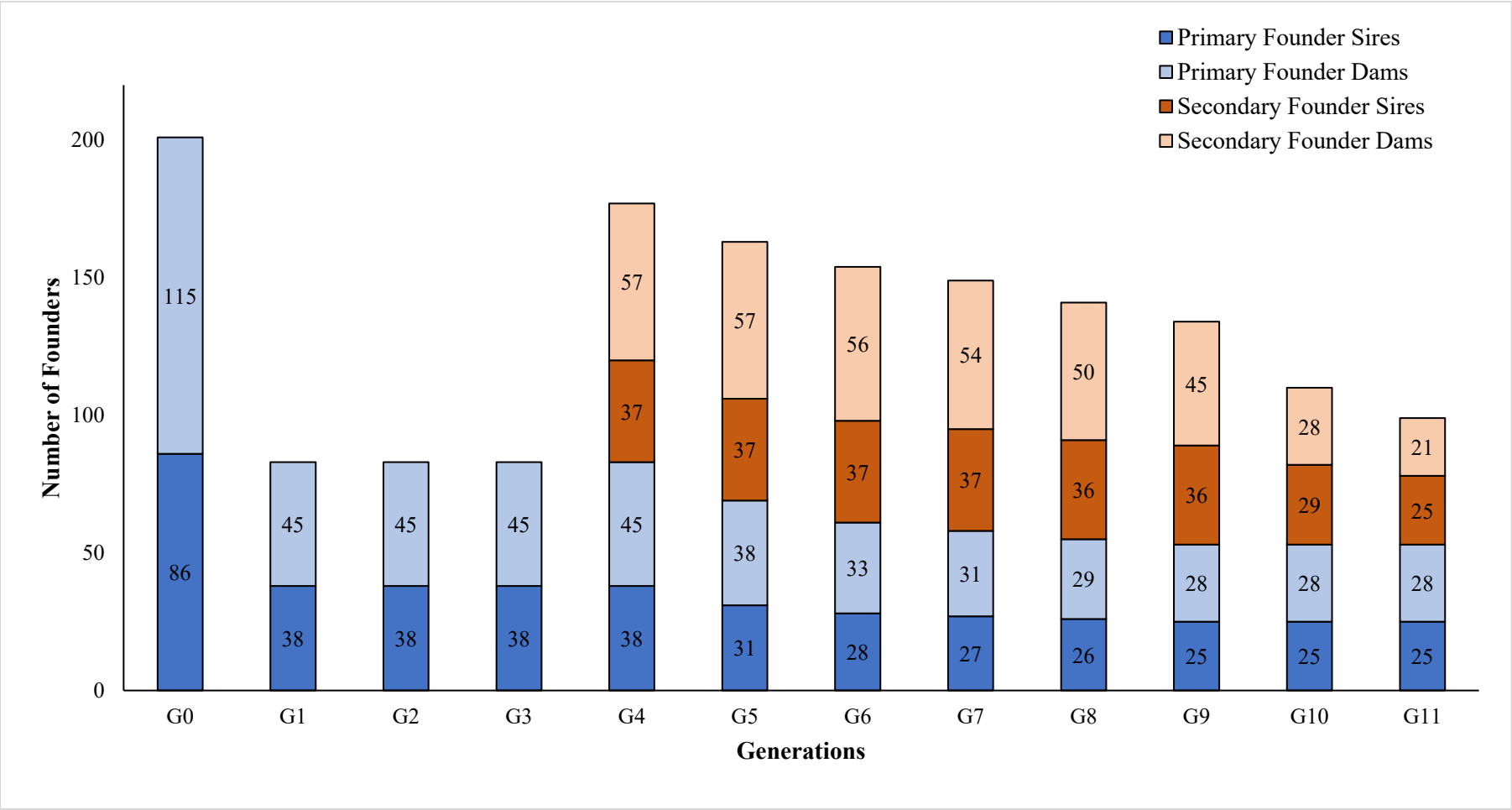
ZHONG, X., LI, Q., KONG, L. & YU, H. 2017. Estimates of Linkage Disequilibrium and Effective Population Size in Wild and Selected Populations of the Pacific Oyster Using Single-nucleotide Polymorphism Markers. *Journal of the World Aquaculture Society,* 48**,** 791-801.

APPENDIX 1

Appendix 1. Classification of pedigree errors. Genealogical and molecular data were used to determine pedigrees. Differences observed between these two assignment methods resulted in pedigree errors, with molecular assignments deemed more accurate than genealogical assignments. Results were categorized into four classes: pedigree agreement, reassigned sires and dams, indiscernible sires and dams, and unknown error status. Of these classes, only reassigned and indiscernible sires and dams were considered to be pedigree errors.

	Could be assigned using genealogical data?	Could be assigned using molecular data?	Did genealogical and molecular assignments agree?	Was it categorized as a pedigree error?	Special molecular assignment conditions
Pedigree Agreement	Yes	Yes	Yes	No	
Reassigned	Yes	Yes	No	Yes	One or both of the following had to be true: • Both pedigree assigned parents had been genotyped • Both parents were molecularly assigned
Indiscernible	Yes	No	No	Yes	Both pedigree assigned parents had been genotyped; however, a molecular parental match was not found within the dataset.
Unknown Error Status	Yes	Yes*	No	No	Both of the following held true: • No, or only 1 parent, had been genotyped • Only one parent could be reassigned

APPENDIX 2



Appendix 2. Number of founder genomes identified in the Abbassa Strain over 11 generations. Dark blue bars denote the number of original AS sire founders, light blue denote the number of original AS dam founders, dark orange bars denote sires introduced to the AS during generation 4, and light orange bars denote dams introduced to the AS during generation 4.

APPENDIX 3

Appendix 3. The number of offspring per founder as well as their overall genome contribution to generations 1, 5, and 11 are provided. Only 83 of the 201 primary founders were identified in pedigree records. Founders highlighted in red text provided 0 or negligible genetic contribution to generation 11 (less than 0.001, or less than 0.1% genome contribution). Cells in bolded blue text indicate those secondary founders introduced in generation 5. Unknown founder denotes those genetic contributions which could not be assigned due to incomplete pedigrees.

Founder	Sex	Number of Offspring	G1	G5	G11
10690	M	15	0.039	0.046	0.060
10180	M	10	0.028	0.034	0.047
10210	M	4	0.014	0.025	0.042
10700	M	1	0.007	0.024	0.038
10701	F	1	0.007	0.024	0.038
10211	F	3	0.012	0.021	0.036
10181	F	6	0.019	0.024	0.035
10692	F	3	0.012	0.024	0.035
10220	M	5	0.016	0.021	0.033
10221	F	2	0.009	0.020	0.033
10920	M	7	0.021	0.026	0.030
10921	F	7	0.021	0.026	0.030
10250	F	2	0.028	0.024	0.027
10252	M	1	0.028	0.024	0.027
10691	F	1	0.032	0.023	0.025
10230	M	10	0.014	0.019	0.024
10231	F	10	0.014	0.019	0.024
10170	F	12	0.014	0.018	0.019
10171	M	4	0.014	0.018	0.019
10470	F	4	0.016	0.016	0.019
10471	M	4	0.016	0.016	0.019
10400	F	4	0.007	0.013	0.018
10401	M	5	0.007	0.013	0.018
10570	F	5	0.016	0.014	0.017
10571	M	1	0.016	0.014	0.017
10520	F	1	0.012	0.012	0.016
10521	M	5	0.009	0.010	0.016
10130	F	5	0.007	0.011	0.013
10131	M	3	0.007	0.011	0.013
10182	F	4	0.014	0.010	0.011
10960	M	6	0.019	0.007	0.011
10961	F	6	0.019	0.007	0.011
10590	M	4	0.014	0.005	0.010

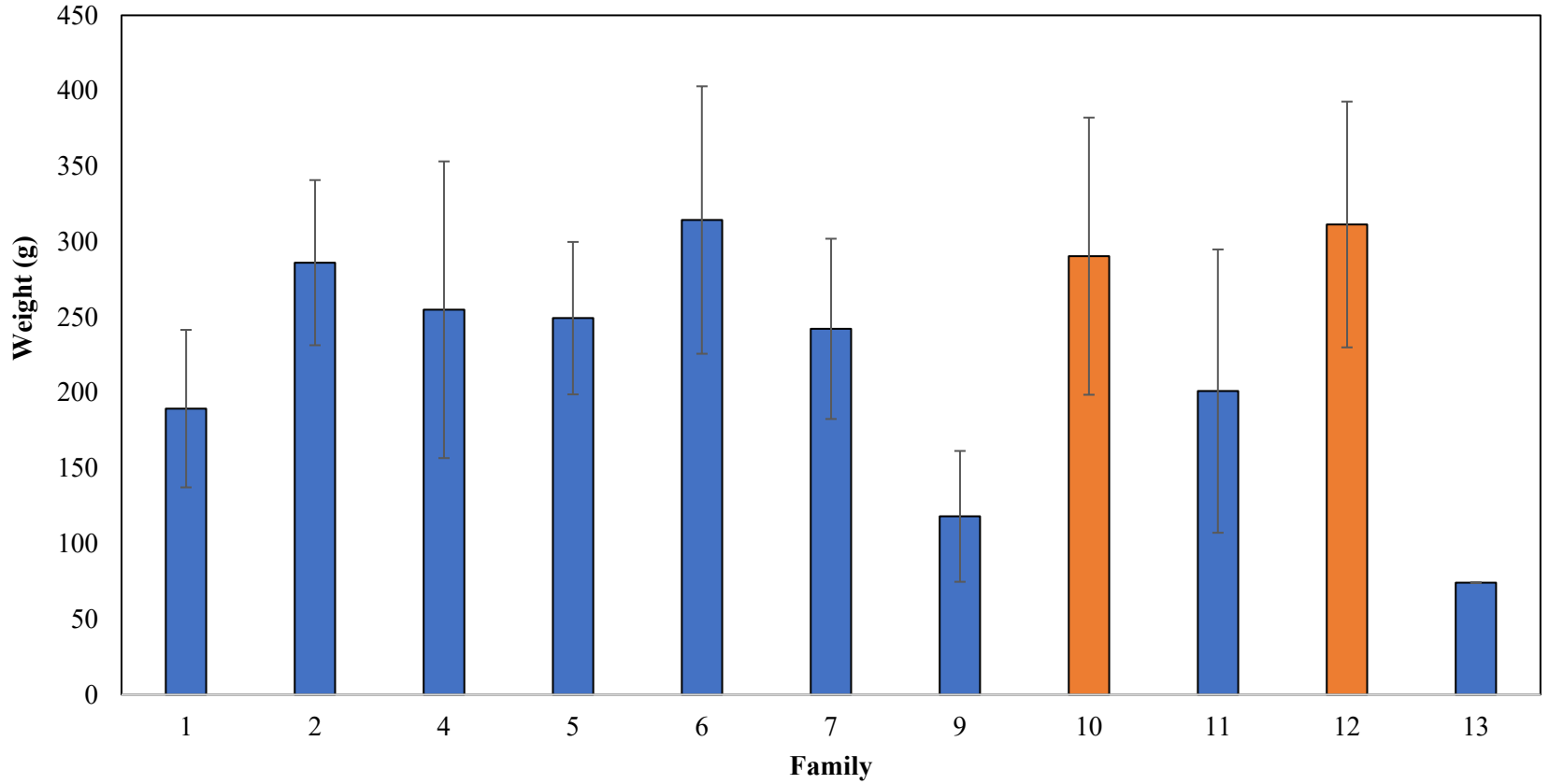
10591	F	4	0.014	0.005	0.010
10910	M	3	0.012	0.004	0.006
10912	F	3	0.012	0.004	0.006
10260	M	5	0.016	0.004	0.006
10261	F	5	0.016	0.004	0.006
10212	F	1	0.007	0.004	0.006
10200	M	1	0.007	0.002	0.006
10201	F	1	0.007	0.002	0.006
10460	M	2	0.009	0.009	0.004
10461	F	2	0.009	0.009	0.004
10380	M	2	0.009	0.003	0.004
10381	F	2	0.009	0.003	0.004
10600	M	3	0.012	0.003	0.004
10601	F	2	0.009	0.003	0.004
10360	M	1	0.007	0.004	0.002
10361	F	1	0.007	0.004	0.002
10890	M	3	0.012	0.004	0.001
10891	F	3	0.012	0.004	0.001
10100	M	2	0.009	0.003	0.001
10102	F	2	0.009	0.003	0.001
10522	F	1	0.007	0.002	0.000
10670	M	7	0.021	0.002	0.000
10672	F	4	0.014	0.001	0.000
10300	M	4	0.014	0.001	0.000
10301	F	4	0.014	0.001	0.000
10340	M	1	0.007	0.001	0.000
10341	F	1	0.007	0.001	0.000
10671	F	3	0.012	0.001	0.000
10222	F	3	0.012	0.001	0.000
10120	M	2	0.009	0.001	0.000
10121	F	2	0.009	0.001	0.000
10280	M	2	0.009	0.001	0.000
10281	F	2	0.009	0.001	0.000
10350	M	2	0.009	0.001	0.000
10352	F	2	0.009	0.001	0.000
10602	F	1	0.007	0.001	0.000
10500	M	2	0.009	0.000	0.000
10501	F	2	0.009	0.000	0.000
10510	M	1	0.007	0.000	0.000
10512	F	1	0.007	0.000	0.000
10450	M	1	0.007	0.000	0.000
10451	F	1	0.007	0.000	0.000
10150	M	1	0.007	0.000	0.000

10151	F	1	0.007	0.000	0.000
10330	M	1	0.007	0.000	0.000
10331	F	1	0.007	0.000	0.000
10480	M	1	0.007	0.000	0.000
10481	F	1	0.007	0.000	0.000
10850	M	1	0.007	0.000	0.000
10851	F	1	0.007	0.000	0.000
4004023	F	20	0.000	0.003	0.015
4900006	M	38	0.000	0.005	0.002
4910006	F	19	0.000	0.002	0.002
4900019	M	42	0.000	0.005	0.002
4006822	M	50	0.000	0.006	0.002
4900007	M	49	0.000	0.006	0.002
4910039	F	23	0.000	0.003	0.002
4900027	M	41	0.000	0.005	0.001
4900010	M	47	0.000	0.006	0.001
4900024	M	47	0.000	0.006	0.001
4900023	M	40	0.000	0.005	0.001
4900012	M	44	0.000	0.005	0.001
4900018	M	44	0.000	0.005	0.001
4900011	M	32	0.000	0.004	0.001
4900005	M	41	0.000	0.005	0.001
4900003	M	26	0.000	0.003	0.001
4910003	F	26	0.000	0.003	0.001
4013907	M	48	0.000	0.006	0.001
4910047	F	21	0.000	0.003	0.001
4900004	M	46	0.000	0.005	0.001
4900002	M	45	0.000	0.005	0.001
4900020	M	15	0.000	0.002	0.001
4910040	F	15	0.000	0.002	0.001
4900026	M	33	0.000	0.004	0.001
4910007	F	27	0.000	0.003	0.001
4910044	F	25	0.000	0.003	0.001
4910017	F	22	0.000	0.003	0.001
4910010	F	24	0.000	0.003	0.001
4900028	M	41	0.000	0.005	0.001
4910021	F	18	0.000	0.002	0.001
4900001	M	21	0.000	0.003	0.001
4910001	F	21	0.000	0.003	0.001
4900037	M	19	0.000	0.002	0.001
4910067	F	19	0.000	0.002	0.001
4910043	F	20	0.000	0.003	0.001
4910057	F	20	0.000	0.003	0.001

4900016	M	34	0.000	0.004	0.001
4900030	M	22	0.000	0.003	0.001
4910050	F	22	0.000	0.003	0.001
4910028	F	21	0.000	0.003	0.001
4910015	F	19	0.000	0.002	0.001
4900008	M	30	0.000	0.004	0.001
4900021	M	30	0.000	0.004	0.001
4910004	F	29	0.000	0.004	0.001
4910022	F	18	0.000	0.002	0.001
4910046	F	18	0.000	0.002	0.001
4900029	M	26	0.000	0.003	0.000
4910032	F	11	0.000	0.003	0.000
4910049	F	26	0.000	0.003	0.000
4900014	M	25	0.000	0.003	0.000
4900017	M	25	0.000	0.003	0.000
4900013	M	24	0.000	0.003	0.000
4910023	F	24	0.000	0.003	0.000
4910002	F	23	0.000	0.003	0.000
4910020	F	23	0.000	0.003	0.000
4910026	F	23	0.000	0.003	0.000
4910038	F	23	0.000	0.003	0.000
4910005	F	22	0.000	0.003	0.000
4910012	F	22	0.000	0.003	0.000
4910024	F	22	0.000	0.003	0.000
4910054	F	22	0.000	0.003	0.000
4900022	M	21	0.000	0.003	0.000
4900025	M	21	0.000	0.003	0.000
4900035	M	21	0.000	0.003	0.000
4910041	F	21	0.000	0.003	0.000
4910052	F	21	0.000	0.003	0.000
4910055	F	21	0.000	0.003	0.000
4910058	F	21	0.000	0.003	0.000
4910065	F	21	0.000	0.003	0.000
4910048	F	20	0.000	0.003	0.000
4910053	F	20	0.000	0.003	0.000
4910016	F	19	0.000	0.002	0.000
4910027	F	19	0.000	0.002	0.000
4910029	F	19	0.000	0.002	0.000
4900009	M	18	0.000	0.002	0.000
4900015	M	18	0.000	0.002	0.000
4900034	M	18	0.000	0.002	0.000
4900036	M	18	0.000	0.002	0.000
4910019	F	18	0.000	0.002	0.000

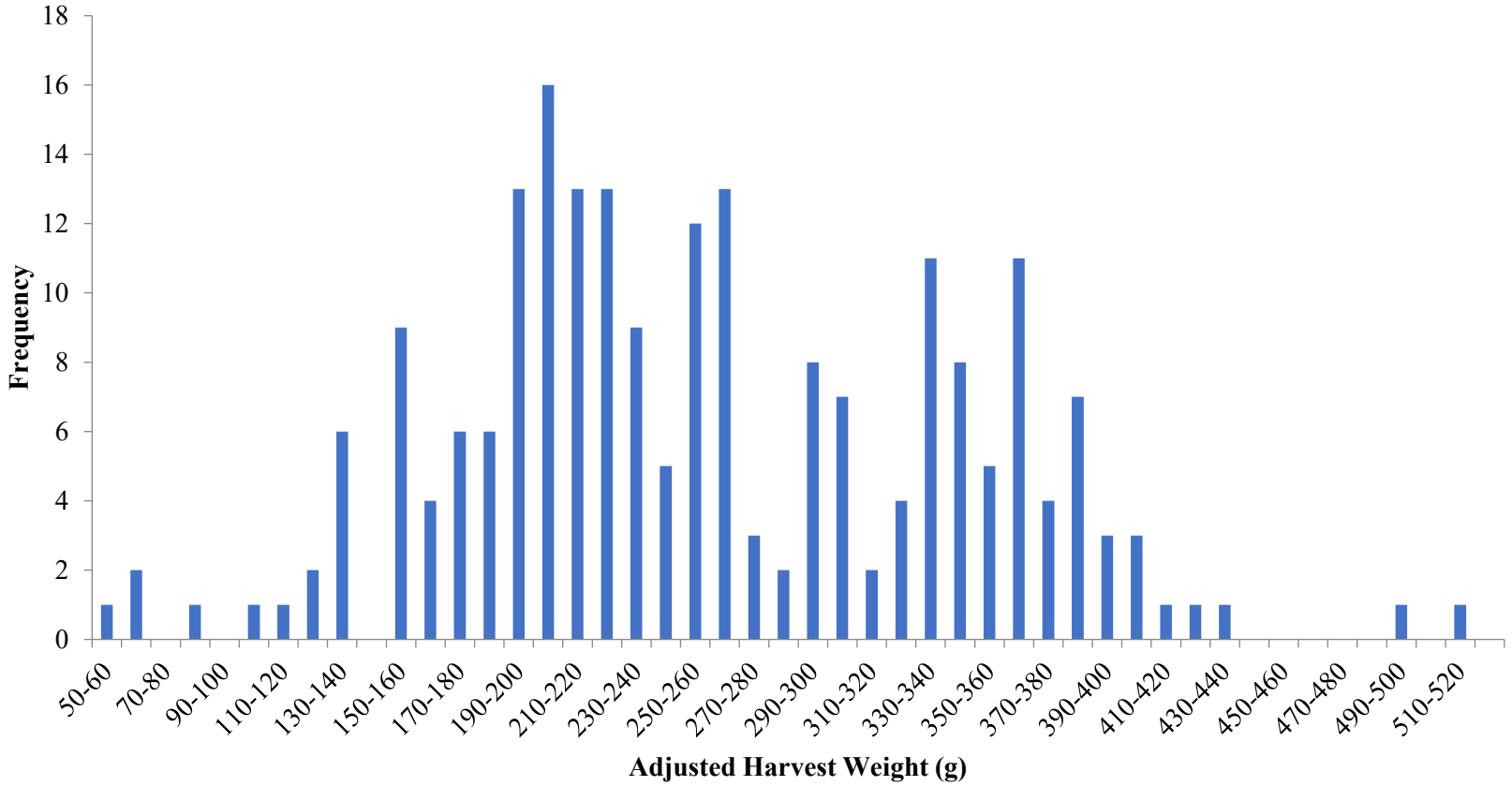
4910064	F	18	0.000	0.002	0.000
4910066	F	18	0.000	0.002	0.000
4910008	F	17	0.000	0.002	0.000
4910014	F	17	0.000	0.002	0.000
4910056	F	15	0.000	0.002	0.000
4910031	F	14	0.000	0.002	0.000
4910018	F	13	0.000	0.002	0.000
4900032	M	11	0.000	0.001	0.000
4910025	F	11	0.000	0.001	0.000
4910036	F	11	0.000	0.001	0.000
4910062	F	11	0.000	0.001	0.000
4910051	F	9	0.000	0.001	0.000
4910035	F	7	0.000	0.001	0.000
4910037	F	6	0.000	0.001	0.000
4910034	F	3	0.000	0.001	0.000
Unknown			0.000	0.001	0.007

APPENDIX 4



Appendix 4. Average final weight (g) adjusted by the average number of days from spawn to harvest for all families with ≥5 offspring of the Abbassa Strain of Nile tilapia, with error bars representing one standard deviation. Not all offspring had phenotypic data available; therefore, some family averages are based on fewer individuals with Family 13 only having a single offspring with phenotypic data. Families in orange are those families used for QTL, GWAS, and linkage mapping analysis, with blue signifying families used only for GWAS and linkage mapping.

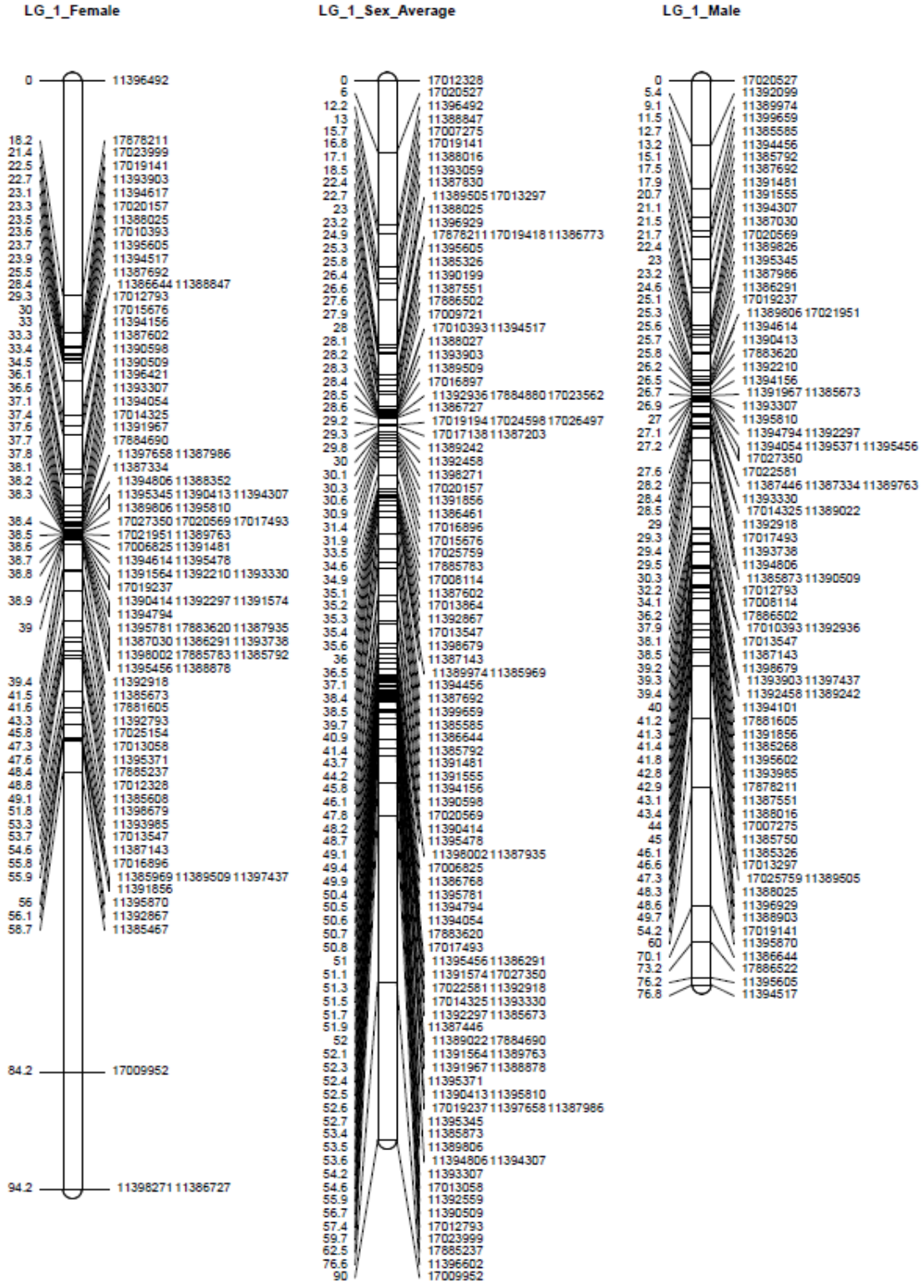
APPENDIX 5

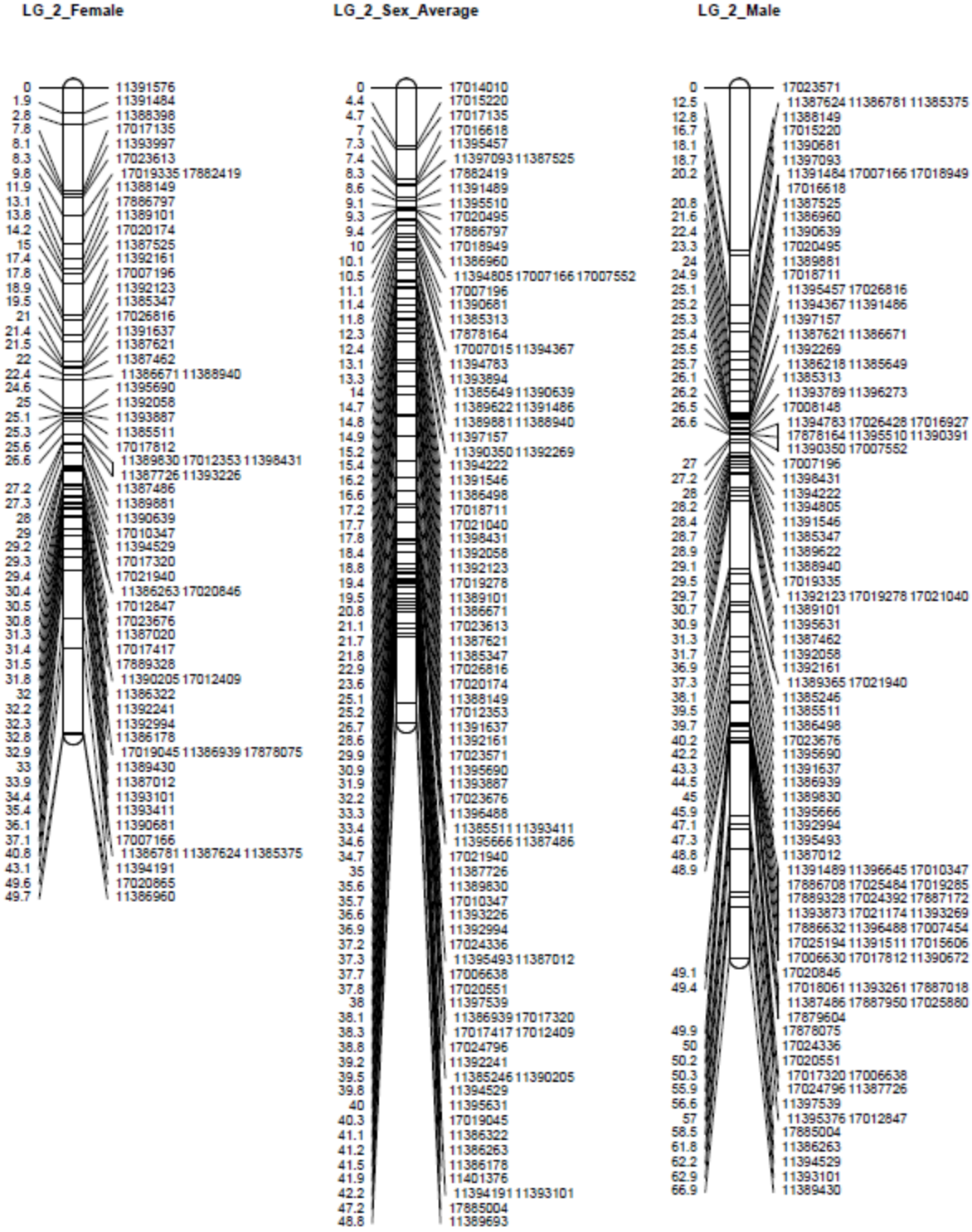


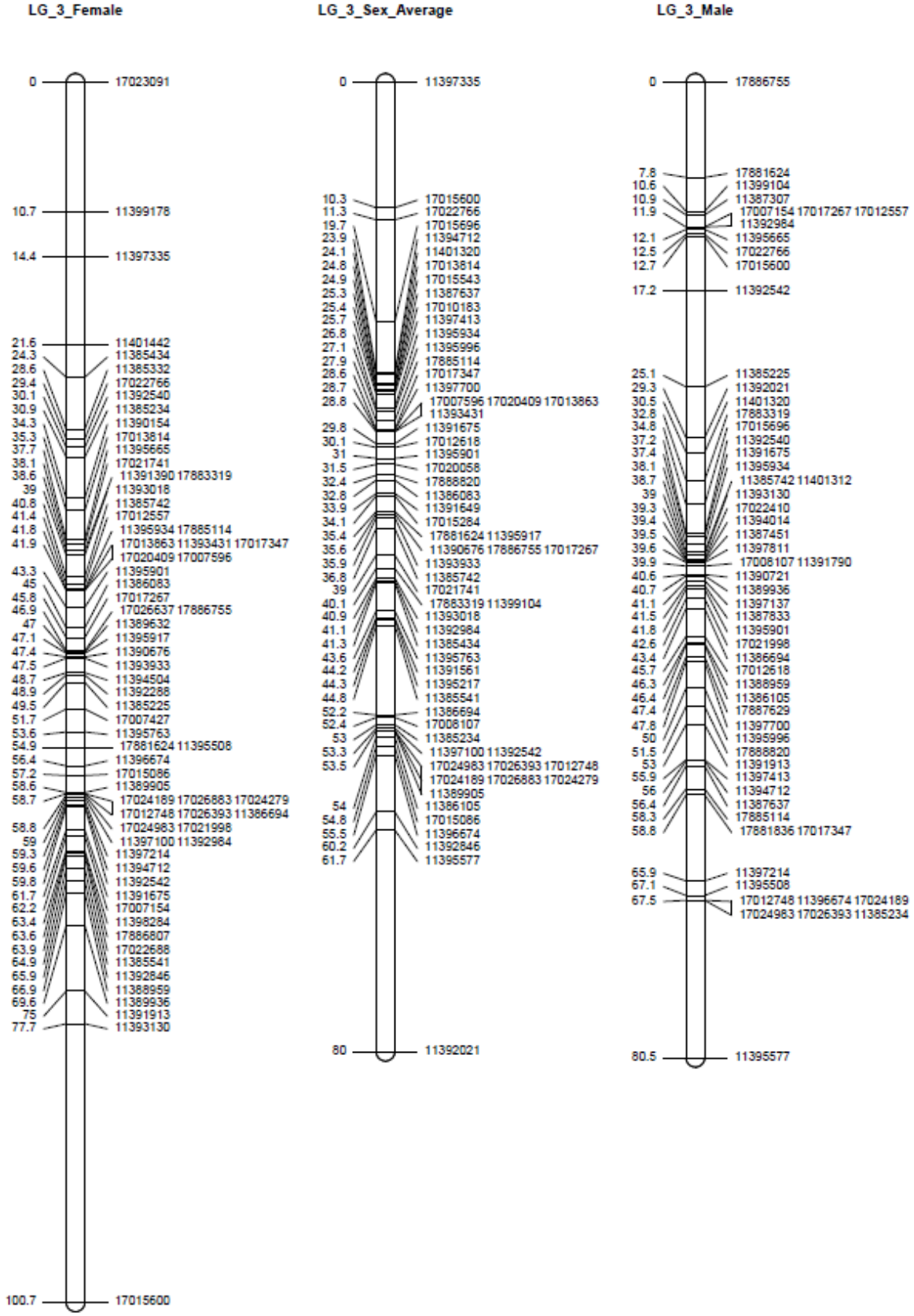
Appendix 5. Weight classes observed in Generation 10 of the Abbassa Strain in 10g bins. Harvest weight (g) were adjusted by the average number of days from spawn to harvest.

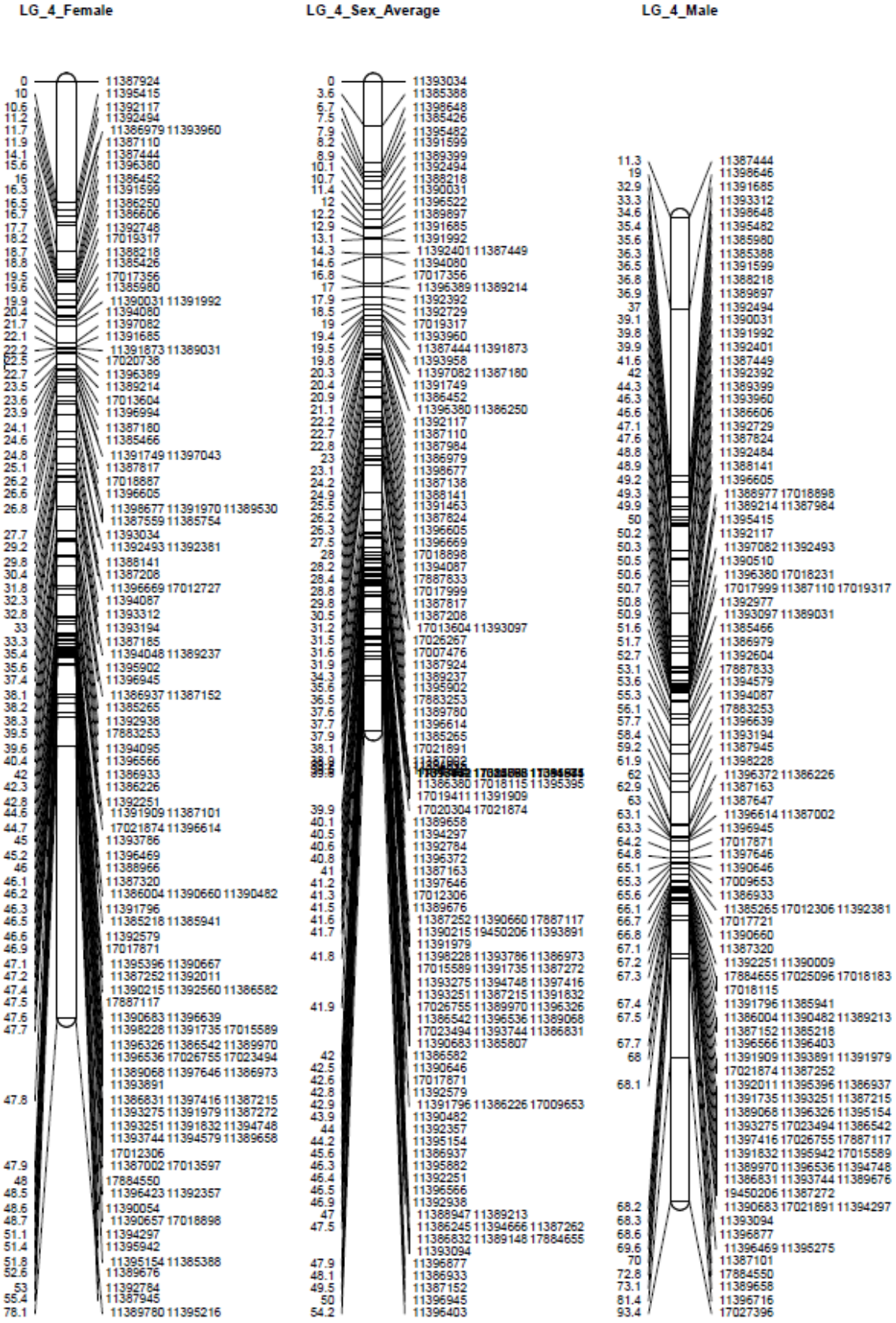
APPENDIX 6

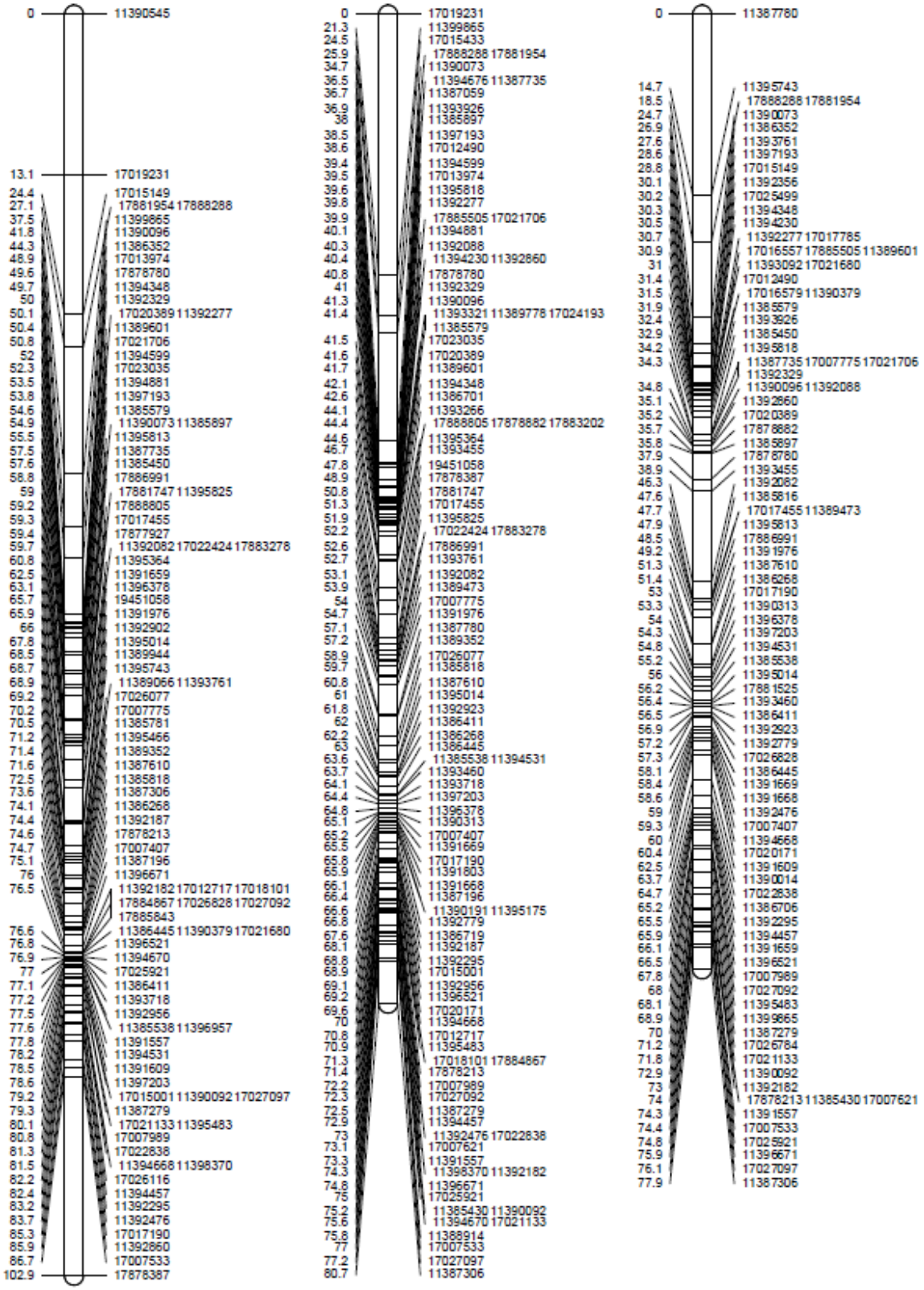
Appendix 6. Linkage groups based on sex Average, female, and male maps.

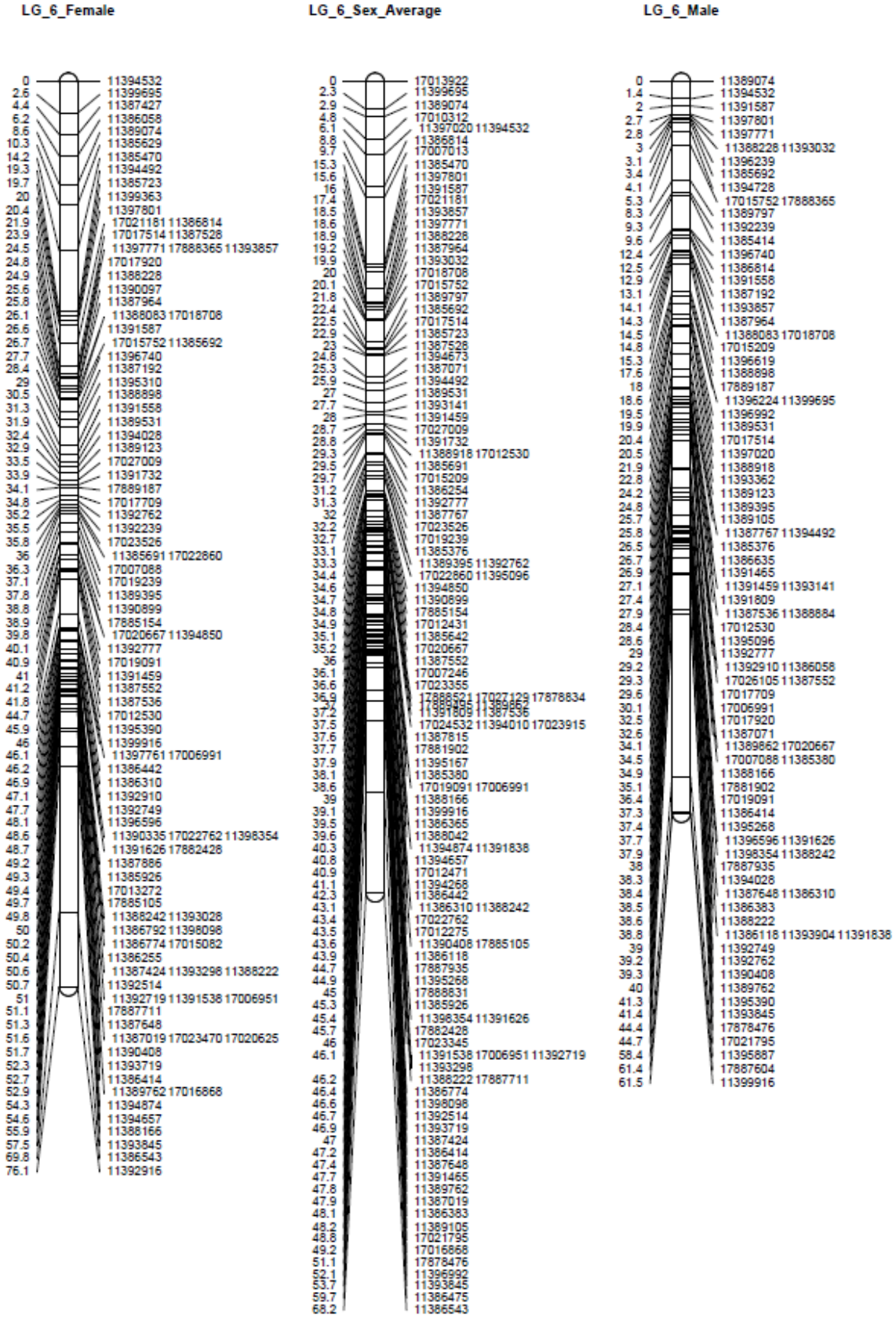


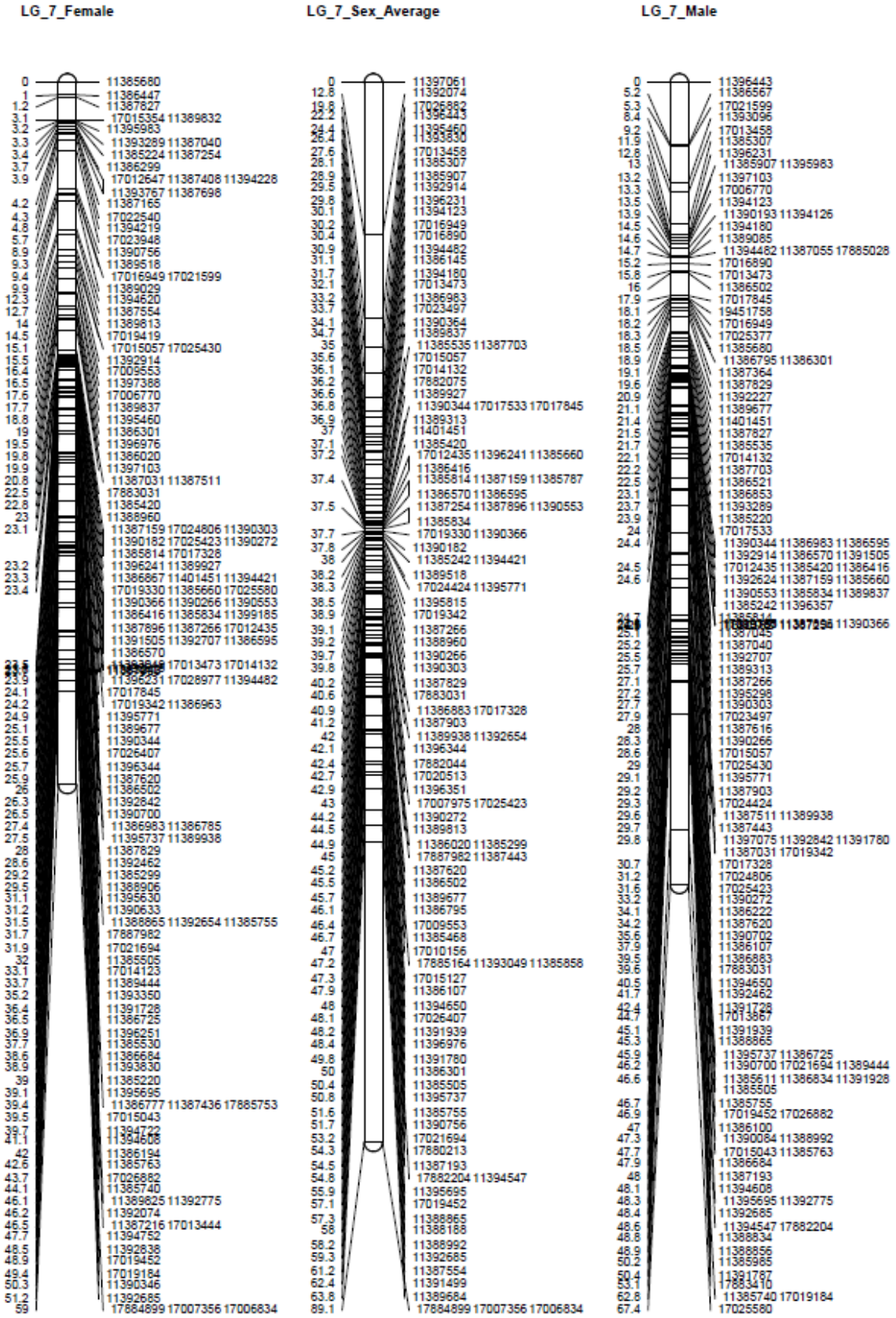

LG_2_Female
0 11391576
1.9 11391484
2.6 11388398
7.8 17017135
8.1 11393997
8.3 17023613
9.6 17019335 17882419
11.9 11388149
13.1 17886797
13.8 11389101
14.2 17020174
15 11387525
17.4 11392161
17.8 17007196
18.9 11392123
19.5 11385347
21 17026816
21.4 11391637
21.5 11387621
22 11387462
22.4 11386671 11388940
24.6 11395690
25 11392058
25.1 11393887
25.3 11385511
25.6 17017812
26.6 11389830 17012353 11398431
11387726 11393226
27.2 11387486
27.3 11389881
28 11390639
29 17010347
29.2 11394529
29.3 17017320
29.4 17021940
30.4 11386263 17020846
30.5 17012847
30.8 17023676
31.3 11387020
31.4 17017417
31.5 17889328
31.8 11390205 17012409
32 11386322
32.2 11392241
32.3 11392994
32.8 11386178
32.9 17019045 11386939 17878075
33 11389430
33.9 11387012
34.4 11393101
35.4 11393411
36.1 11390681
37.1 17007166
40.8 11386781 11387624 11385375
43.1 11394191
49.6 17020865
49.7 11386960
LG_2_Sex_Average
0 17014010
4.4 17015220
4.7 17017135
7 17016618
7.3 11395457
7.4 11397093 11387525
8.3 17882419
8.6 11391489
9.1 11395510
9.3 17020495
9.4 17886797
10 17018949
10.1 11386960
10.5 11394805 17007166 17007552
11.1 17007196
11.4 11390681
11.8 11385313
12.3 17878164
12.4 17007015 11394367
13.1 11394783
13.3 11393894
14 11385649 11390639
14.7 11389622 11391486
14.8 11389881 11388940
14.9 11397157
15.2 11390350 11392269
15.4 11394222
16.2 11391546
16.6 11386498
17.2 17018711
17.7 17021040
17.8 11398431
18.4 11392058
18.8 11392123
19.4 17019278
19.5 11389101
20.8 11386671
21.1 17023613
21.7 11387621
21.8 11385347
22.9 17026816
23.6 17020174
25.1 11388149
25.2 17012353
26.7 11391637
28.6 11392161
29.9 17023571
30.9 11395690
31.9 11393887
32.2 17023676
33.3 11396488
33.4 11385511 11393411
34.6 11395666 11387486
34.7 17021940
35 11387726
35.6 11389630
35.7 17010347
36.6 11393226
36.9 11392994
37.2 17024336
37.3 11395493 11387012
37.7 17006638
37.8 17020551
38 11397539
38.1 11386939 17017320
38.3 17017417 17012409
38.8 17024796
39.2 11392241
39.5 11385246 11390205
39.8 11394529
40 11395631
40.3 17019045
41.1 11386322
41.2 11386263
41.5 11386178
41.9 11401376
42.2 11394191 11393101
47.2 17885004
48.8 11389693
LG_2_Male
0 17023571
12.5 11387624 11386781 11385375
12.8 11388149
16.7 17015220
18.1 11390681
18.7 11397093
20.2 11391484 17007166 17018949
17016618
20.8 11387525
21.6 11386960
22.4 11390639
23.3 17020495
24 11389881
24.9 17018711
25.1 11395457 17026816
25.2 11394367 11391486
25.3 11397157
25.4 11387621 11386671
25.5 11392269
25.7 11386218 11385649
26.1 11385313
26.2 11393789 11396273
26.5 17008148
26.6 11394783 17026428 17016927
17878164 11395510 11390391
11390350 17007552
27 17007196
27.2 11398431
28 11394222
28.2 11394805
28.4 11391546
28.7 11385347
28.9 11389622
29.1 11388940
29.5 17019335
29.7 11392123 17019278 17021040
30.7 11389101
30.9 11395631
31.3 11387462
31.7 11392058
36.9 11392161
37.3 11389365 17021940
38.1 11385246
39.5 11385511
39.7 11386498
40.2 17023676
42.2 11395690
43.3 11391637
44.5 11386939
45 11389830
45.9 11395666
47.1 11392994
47.3 11395493
48.8 11387012
48.9 11391489 11396645 17010347
17886708 17025484 17019285
17889328 17024392 17887172
11393873 17021174 11393269
17886632 11396488 17007454
17025194 11391511 17015606
17006630 17017812 11390672
49.1 17020846
49.4 17018061 11393261 17887018
11387486 17887950 17025880
17879604
49.9 17878075
50 17024336
50.2 17020551
50.3 17017320 17006638
55.9 17024796 11387726
56.6 11397539
57 11395376 17012847
58.5 17885004
61.8 11386263
62.2 11394529
62.9 11393101
66.9 11389430

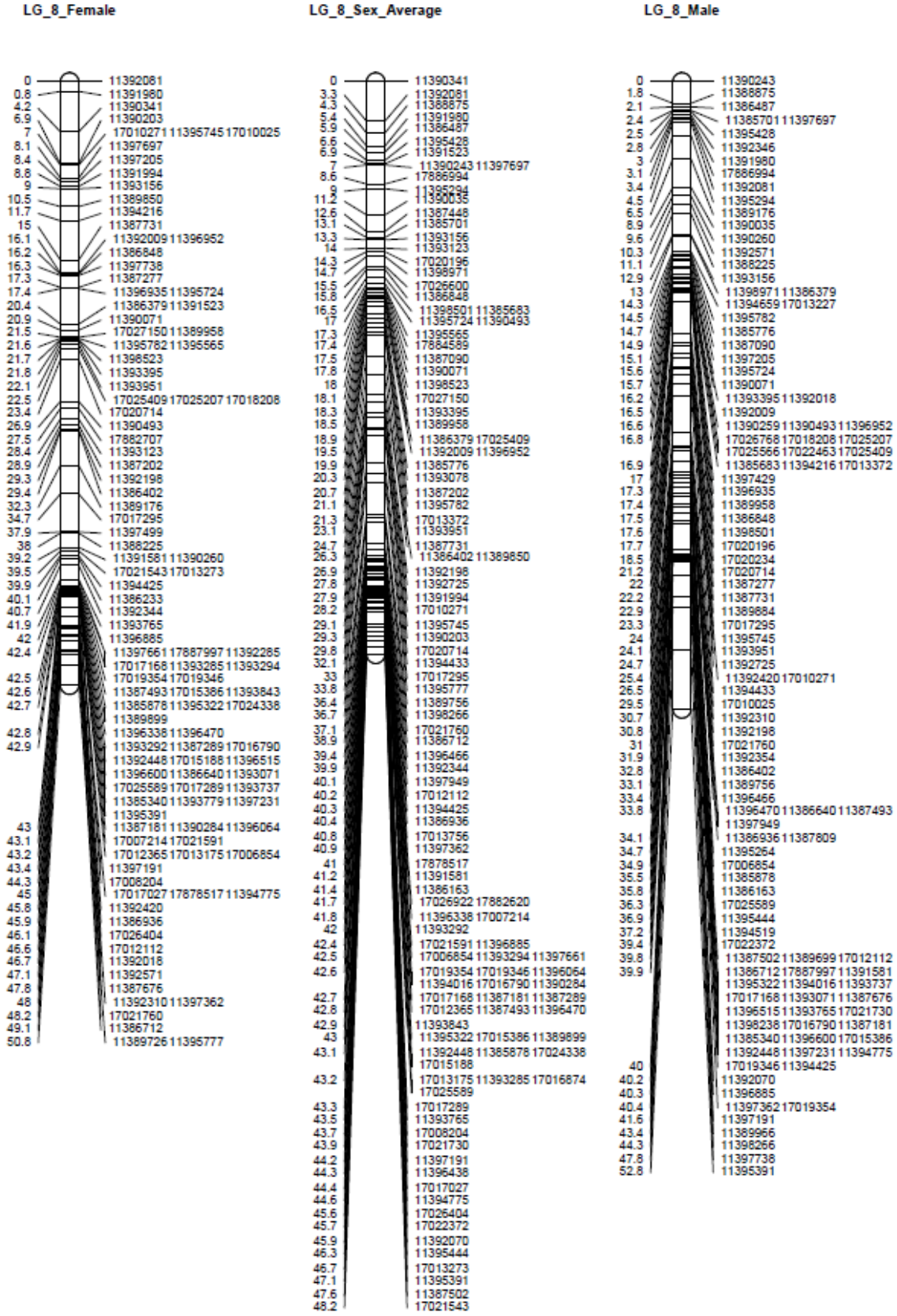

LG_3_Female
0 17023091
10.7 11399178
14.4 11397335
21.6 11401442
24.3 11385434
28.6 11385332
29.4 17022766
30.1 11392540
30.9 11385234
34.3 11390154
35.3 17013814
37.7 11395665
38.1 17021741
38.6 11391390 17883319
39 11393018
40.8 11385742
41.4 17012557
41.8 11395934 17885114
41.9 17013863 11393431 17017347
17020409 17007596
43.3 11395901
45 11386083
45.8 17017267
46.9 17026637 17886755
47 11389632
47.1 11395917
47.4 11390676
47.5 11393933
48.7 11394504
48.9 11392288
49.5 11385225
51.7 17007427
53.6 11395763
54.9 17881624 11395508
56.4 11396674
57.2 17015086
58.6 11389905
58.7 17024189 17026883 17024279
17012748 17026393 11386694
58.8 17024983 17021998
59 11397100 11392984
59.3 11397214
59.6 11394712
59.8 11392542
61.7 11391675
62.2 17007154
63.4 11398284
63.6 17886807
63.9 17022688
64.9 11385541
65.9 11392846
66.9 11388959
69.6 11389936
75 11391913
77.7 11393130
100.7 17015600
LG_3_Sex_Average
0 11397335
10.3 17015600
11.3 17022766
19.7 17015696
23.9 11394712
24.1 11401320
24.8 17013814
24.9 17015543
25.3 11387637
25.4 17010183
25.7 11397413
26.6 11395934
27.1 11395996
27.9 17885114
28.6 17017347
28.7 11397700
28.8 17007596 17020409 17013863
11393431
29.8 11391675
30.1 17012618
31 11395901
31.5 17020058
32.4 17888820
32.8 11386083
33.9 11391649
34.1 17015284
35.4 17881624 11395917
35.6 11390676 17886755 17017267
35.9 11393933
36.8 11385742
39 17021741
40.1 17883319 11399104
40.9 11393018
41.1 11392984
41.3 11385434
43.6 11395763
44.2 11391561
44.3 11395217
44.8 11385541
52.2 11386694
52.4 17008107
53 11385234
53.3 11397100 11392542
53.5 17024983 17026393 17012748
17024189 17026883 17024279
11389905
54 11386105
54.8 17015086
55.5 11396674
60.2 11392846
61.7 11395577
80 11392021
LG_3_Male
0 17886755
7.8 17881624
10.6 11399104
10.9 11387307
11.9 17007154 17017267 17012557
11392984
12.1 11395665
12.5 17022766
12.7 17015600
17.2 11392542
25.1 11385225
29.3 11392021
30.5 11401320
32.8 17883319
34.8 17015696
37.2 11392540
37.4 11391675
38.1 11395934
38.7 11385742 11401312
39 11393130
39.3 17022410
39.4 11394014
39.5 11387451
39.6 11397811
39.9 17008107 11391790
40.6 11390721
40.7 11389936
41.1 11397137
41.5 11387833
41.8 11395901
42.6 17021998
43.4 11386694
45.7 17012618
46.3 11388959
46.4 11386105
47.4 17887629
47.8 11397700
50 11395996
51.5 17888820
53 11391913
55.9 11397413
56 11394712
56.4 11387637
58.3 17885114
58.8 17881836 17017347
65.9 11397214
67.1 11395508
67.5 17012748 11396674 17024189
17024983 17026393 11385234
80.5 11395577

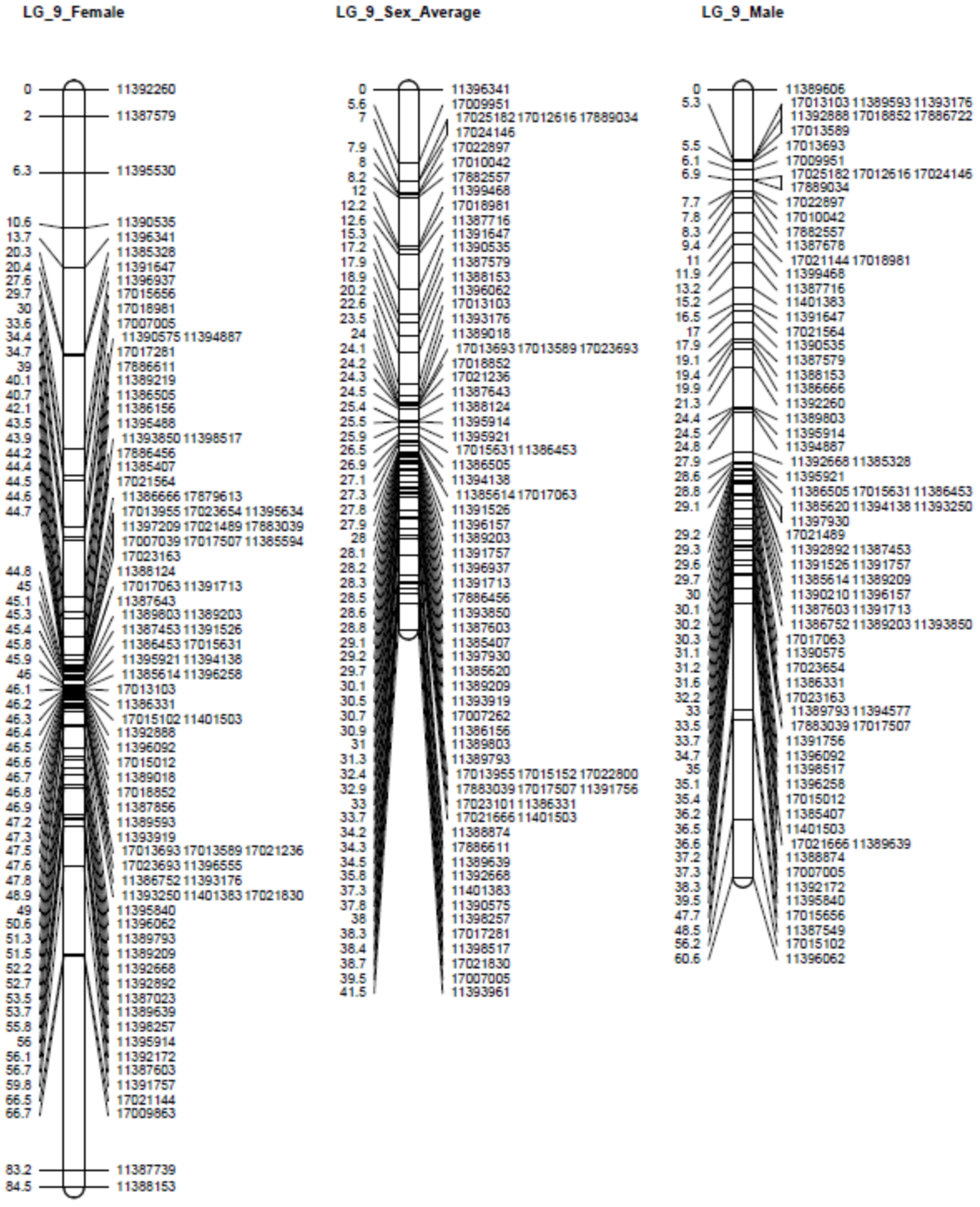

LG_4_Female
LG_4_Sex_Average
LG_4_Male

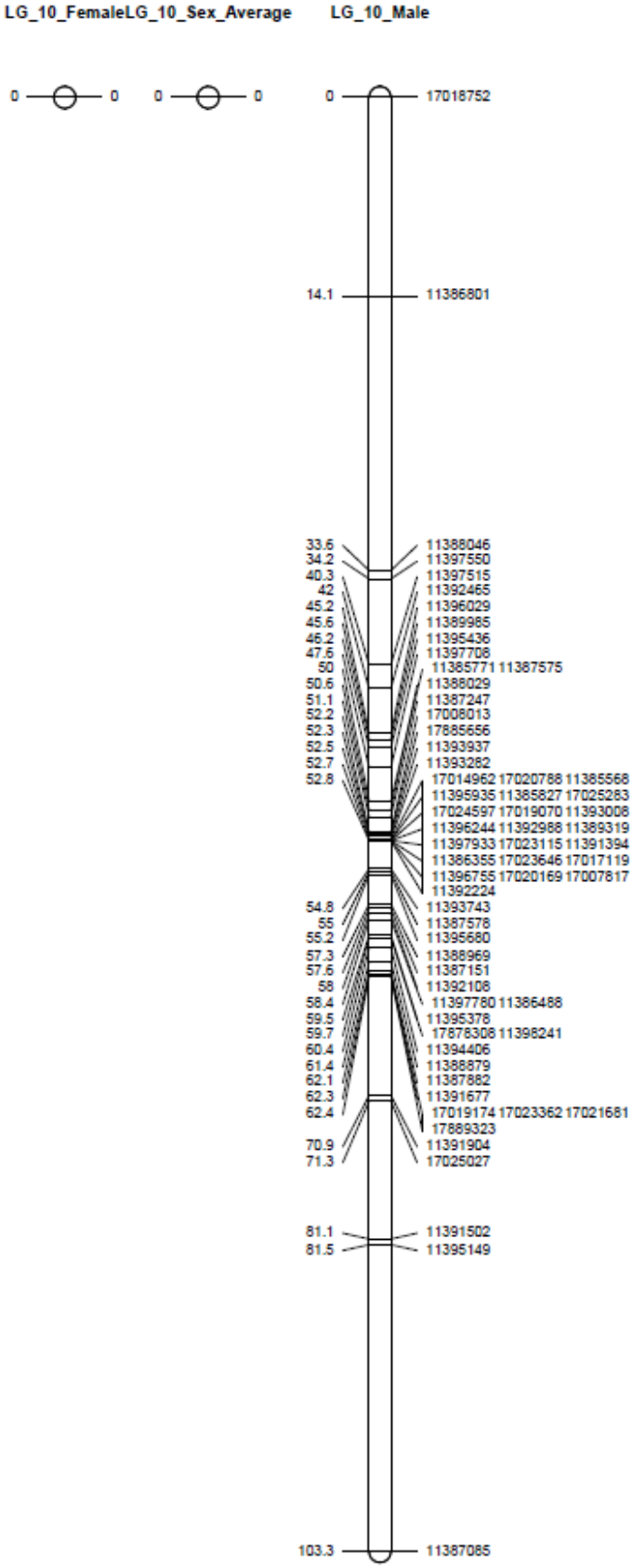

LG_5_Female
0 11390545
13.1 17019231
24.4 17015149
27.1 17881954 17888288
37.5 11399865
41.8 11390096
44.3 11386352
48.9 17013974
49.6 17878780
49.7 11394348
50 11392329
50.1 17020389 11392277
50.4 11389601
50.8 17021706
52 11394599
52.3 17023035
53.5 11394881
53.8 11397193
54.6 11385579
54.9 11390073 11385897
55.5 11395813
57.5 11387735
57.6 11385450
58.8 17886991
59 17881747 11395825
59.2 17888805
59.3 17017455
59.4 17877927
59.7 11392082 17022424 17883278
60.8 11395364
62.5 11391659
63.1 11396378
65.7 19451058
65.9 11391976
66 11392902
67.8 11395014
68.5 11389944
68.7 11395743
68.9 11389066 11393761
69.2 17026077
70.2 17007775
70.5 11385781
71.2 11395466
71.4 11389352
71.6 11387610
72.5 11385818
73.6 11387306
74.1 11386268
74.4 11392187
74.6 17878213
74.7 17007407
75.1 11387196
76 11396671
76.5 11392182 17012717 17018101
17884867 17026828 17027092
17885843
76.6 11386445 11390379 17021680
76.8 11396521
76.9 11394670
77 17025921
77.1 11386411
77.2 11393718
77.5 11392956
77.6 11385538 11396957
77.8 11391557
78.2 11394531
78.5 11391609
78.6 11397203
79.2 17015001 11390092 17027097
79.3 11387279
80.1 17021133 11395483
80.8 17007989
81.3 17022838
81.5 11394668 11398370
82.2 17026116
82.4 11394457
83.2 11392295
83.7 11392476
85.3 17017190
85.9 11392860
86.7 17007533
102.9 17878387
LG_5_Sex_Average
0 17019231
21.3 11399865
24.5 17015433
25.9 17888288 17881954
34.7 11390073
36.5 11394676 11387735
36.7 11387059
36.9 11393926
38 11385897
38.5 11397193
38.6 17012490
39.4 11394599
39.5 17013974
39.6 11395818
39.8 11392277
39.9 17885505 17021706
40.1 11394881
40.3 11392088
40.4 11394230 11392860
40.8 17878780
41 11392329
41.3 11390096
41.4 11393321 11389778 17024193
11385579
41.5 17023035
41.6 17020389
41.7 11389601
42.1 11394348
42.6 11386701
44.1 11393266
44.4 17888805 17878882 17883202
44.6 11395364
46.7 11393455
47.8 19451058
48.9 17878387
50.8 17881747
51.3 17017455
51.9 11395825
52.2 17022424 17883278
52.6 17886991
52.7 11393761
53.1 11392082
53.9 11389473
54 17007775
54.7 11391976
57.1 11387780
57.2 11389352
58.9 17026077
59.7 11385818
60.8 11387610
61 11395014
61.8 11392923
62 11386411
62.2 11386268
63 11386445
63.6 11385538 11394531
63.7 11393460
64.1 11393718
64.4 11397203
64.8 11396378
65.1 11390313
65.2 17007407
65.5 11391669
65.8 17017190
65.9 11391803
66.1 11391668
66.4 11387196
66.6 11390191 11395175
66.8 11392779
67.6 11386719
68.1 11392187
68.8 11392295
68.9 17015001
69.1 11392956
69.2 11396521
69.6 17020171
70 11394668
70.8 17012717
70.9 11395483
71.3 17018101 17884867
71.4 17878213
72.2 17007989
72.3 17027092
72.5 11387279
72.9 11394457
73 11392476 17022838
73.1 17007621
73.3 11391557
74.3 11396370 11392182
74.8 11396671
75 17025921
75.2 11385430 11390092
75.6 11394670 17021133
75.8 11388914
77 17007533
77.2 17027097
80.7 11387306
LG_5_Male
0 11387780
14.7 11395743
18.5 17888288 17881954
24.7 11390073
26.9 11386352
27.6 11393761
28.6 11397193
28.8 17015149
30.1 11392356
30.2 17025499
30.3 11394348
30.5 11394230
30.7 11392277 17017785
30.9 17016557 17885505 11389601
31 11393092 17021680
31.4 17012490
31.5 17016579 11390379
31.9 11385579
32.4 11393926
32.9 11385450
34.2 11395818
34.3 11387735 17007775 17021706
11392329
34.8 11390096 11392088
35.1 11392860
35.2 17020389
35.7 17878882
35.8 11385897
37.9 17878780
38.9 11393455
46.3 11392082
47.6 11385816
47.7 17017455 11389473
47.9 11395813
48.5 17886991
49.2 11391976
51.3 11387610
51.4 11386268
53 17017190
53.3 11390313
54 11396378
54.3 11397203
54.8 11394531
55.2 11385538
56 11395014
56.2 17881525
56.4 11393460
56.5 11386411
56.9 11392923
57.2 11392779
57.3 17026828
58.1 11386445
58.4 11391669
58.6 11391668
59 11392476
59.3 17007407
60 11394668
60.4 17020171
62.5 11391609
63.7 11390014
64.7 17022838
65.2 11386706
65.5 11392295
65.9 11394457
66.1 11391659
66.5 11396521
67.8 17007989
68 17027092
68.1 11395483
68.9 11399865
70 11387279
71.2 17026784
71.8 17021133
72.9 11390092
73 11392182
74 17878213 11385430 17007621
74.3 11391557
74.4 17007533
74.8 17025921
75.9 11396671
76.1 17027097
77.9 11387306

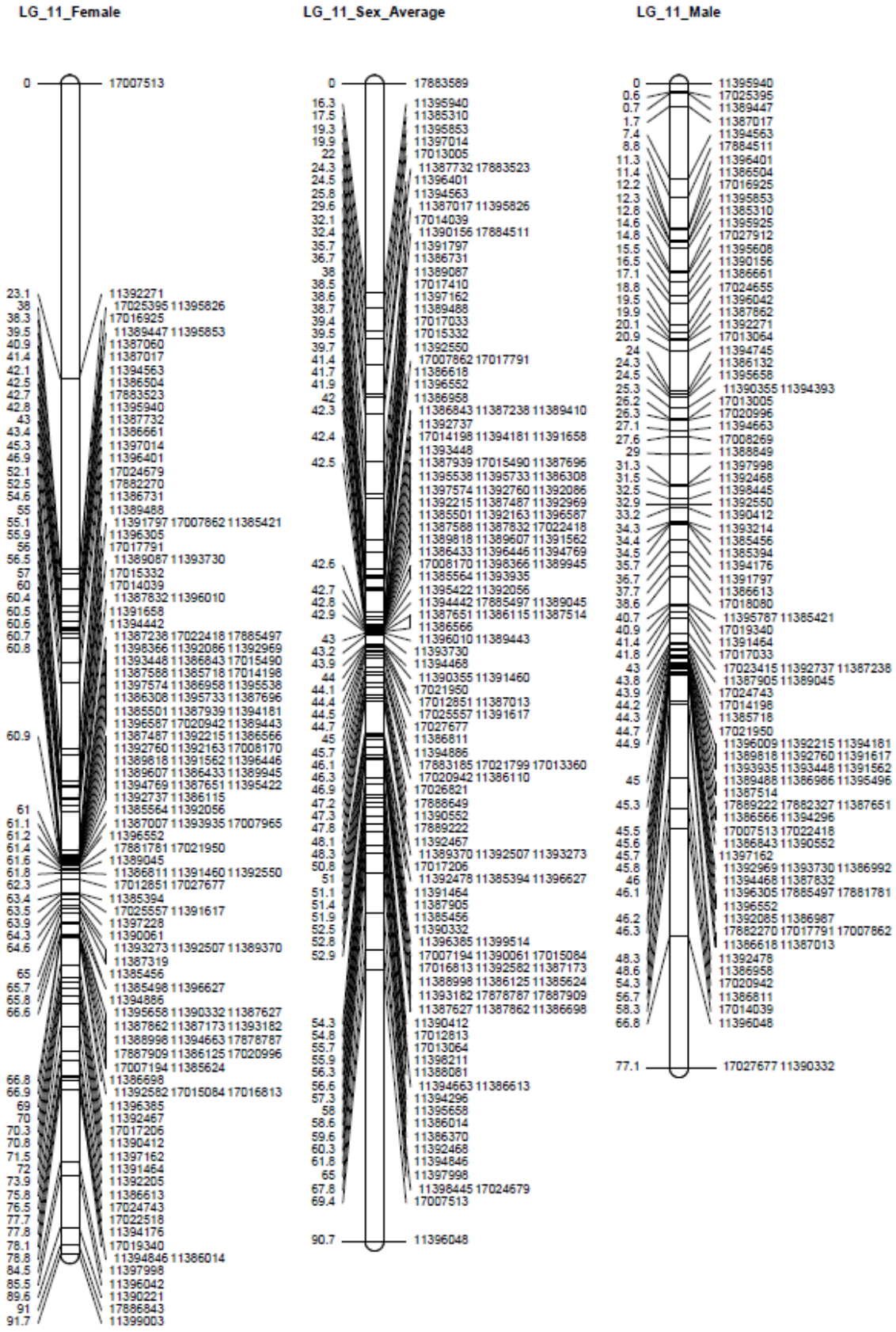

LG_6_Female
0 11394532
2.6 11399695
4.4 11387427
6.2 11386058
8.6 11389074
10.3 11385629
14.2 11385470
19.3 11394492
19.7 11385723
20 11399363
20.4 11397801
21.9 17021181 11386814
23.9 17017514 11387526
24.5 11397771 17888365 11393857
24.8 17017920
24.9 11388228
25.6 11390097
25.8 11387964
26.1 11388083 17018708
26.6 11391587
26.7 17015752 11385692
27.7 11396740
28.4 11387192
29 11395310
30.5 11388898
31.3 11391558
31.9 11389531
32.4 11394028
32.9 11389123
33.5 17027009
33.9 11391732
34.1 17889187
34.8 17017709
35.2 11392762
35.5 11392239
35.8 17023526
36 11385691 17022860
36.3 17007088
37.1 17019239
37.8 11389395
38.8 11390899
38.9 17885154
39.8 17020667 11394850
40.1 11392777
40.9 17019091
41 11391459
41.2 11387552
41.8 11387536
44.7 17012530
45.9 11395390
46 11399916
46.1 11397761 17006991
46.2 11386442
46.9 11386310
47.1 11392910
47.7 11392749
48.1 11396596
48.6 11390335 17022762 11398354
48.7 11391626 17882428
49.2 11387886
49.3 11385926
49.4 17013272
49.7 17885105
49.8 11388242 11393028
50 11386792 11398098
50.2 11386774 17015082
50.4 11386255
50.6 11387424 11393298 11388222
50.7 11392514
51 11392719 11391538 17006951
51.1 17887711
51.3 11387648
51.6 11387019 17023470 17020625
51.7 11390408
52.3 11393719
52.7 11386414
52.9 11389762 17016868
54.3 11394874
54.6 11394657
55.9 11388166
57.5 11393845
69.8 11386543
76.1 11392916
LG_6_Sex_Average
0 17013922
2.3 11399695
2.9 11389074
4.8 17010312
6.1 11397020 11394532
8.8 11386814
9.7 17007013
15.3 11385470
15.6 11397801
16 11391587
17.4 17021181
18.5 11393857
18.6 11397771
18.9 11388228
19.2 11387964
19.9 11393032
20 17018708
20.1 17015752
21.8 11389797
22.4 11385692
22.5 17017514
22.9 11385723
23 11387528
24.8 11394673
25.3 11387071
25.9 11394492
27 11389531
27.7 11393141
28 11391459
28.7 17027009
28.8 11391732
29.3 11388918 17012530
29.5 11385691
29.7 17015209
31.2 11386254
31.3 11392777
32 11387767
32.2 17023526
32.7 17019239
33.1 11385376
33.3 11389395 11392762
34.4 17022860 11395096
34.6 11394850
34.7 11390899
34.8 17885154
34.9 17012431
35.1 11385642
35.2 17020667
36 11387552
36.1 17007246
36.6 17023355
36.9 17888521 17027129 17878834
37 11389405 11389662
37.2 11391609 11387536
37.5 17024532 11394010 17023915
37.6 11387815
37.7 17881902
37.9 11395167
38.1 11385380
38.6 17019091 17006991
39 11388166
39.1 11399916
39.5 11386365
39.6 11388042
40.3 11394874 11391838
40.8 11394657
40.9 17012471
41.1 11394268
42.3 11386442
43.1 11386310 11388242
43.4 17022762
43.5 17012275
43.6 11390408 17885105
43.9 11386118
44.7 17887935
44.9 11395268
45 17888831
45.3 11385926
45.4 11398354 11391626
45.7 17882428
46 17023345
46.1 11391538 17006951 11392719
11393298
46.2 11388222 17887711
46.4 11386774
46.6 11398098
46.7 11392514
46.9 11393719
47 11387424
47.2 11386414
47.4 11387648
47.7 11391465
47.8 11389762
47.9 11387019
48.1 11386383
48.2 11389105
48.8 17021795
49.2 17016868
51.1 17878476
52.1 11396992
53.7 11393845
59.7 11386475
68.2 11386543
LG_6_Male
0 11389074
1.4 11394532
2 11391587
2.7 11397801
2.8 11397771
3 11388228 11393032
3.1 11396239
3.4 11385692
4.1 11394728
5.3 17015752 17888365
8.3 11389797
9.3 11392239
9.6 11385414
12.4 11396740
12.5 11386814
12.9 11391558
13.1 11387192
14.1 11393857
14.3 11387964
14.5 11388083 17018708
14.8 17015209
15.3 11396619
17.6 11388898
18 17889187
18.6 11396224 11399695
19.5 11396992
19.9 11389531
20.4 17017514
20.5 11397020
21.9 11388918
22.8 11393362
24.2 11389123
24.8 11389395
25.7 11389105
25.8 11387767 11394492
26.5 11385376
26.7 11386635
26.9 11391465
27.1 11391459 11393141
27.4 11391809
27.9 11387536 11388884
28.4 17012530
28.6 11395096
29 11392777
29.2 11392910 11386058
29.3 17026105 11387552
29.6 17017709
30.1 17006991
32.5 17017920
32.6 11387071
34.1 11389662 17020667
34.5 17007088 11385380
34.9 11388166
35.1 17881902
36.4 17019091
37.3 11386414
37.4 11395268
37.7 11396596 11391626
37.9 11398354 11388242
38 17887935
38.3 11394028
38.4 11387648 11386310
38.5 11386383
38.6 11388222
38.8 11386118 11393904 11391838
39 11392749
39.2 11392762
39.3 11390408
40 11389762
41.3 11395390
41.4 11393845
44.4 17878476
44.7 17021795
58.4 11395887
61.4 17887604
61.5 11399916

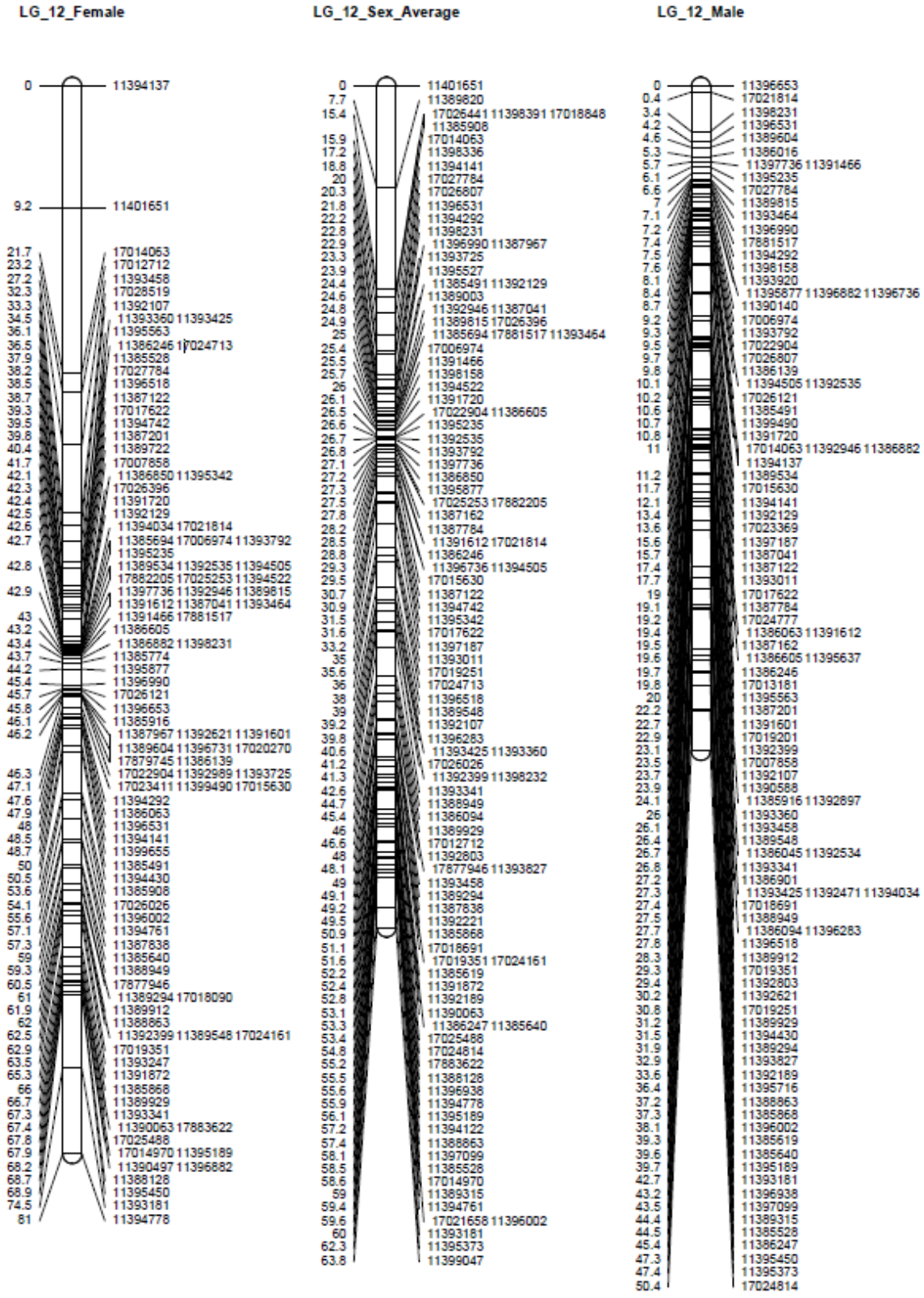

LG_7_Female
LG_7_Sex_Average
LG_7_Male

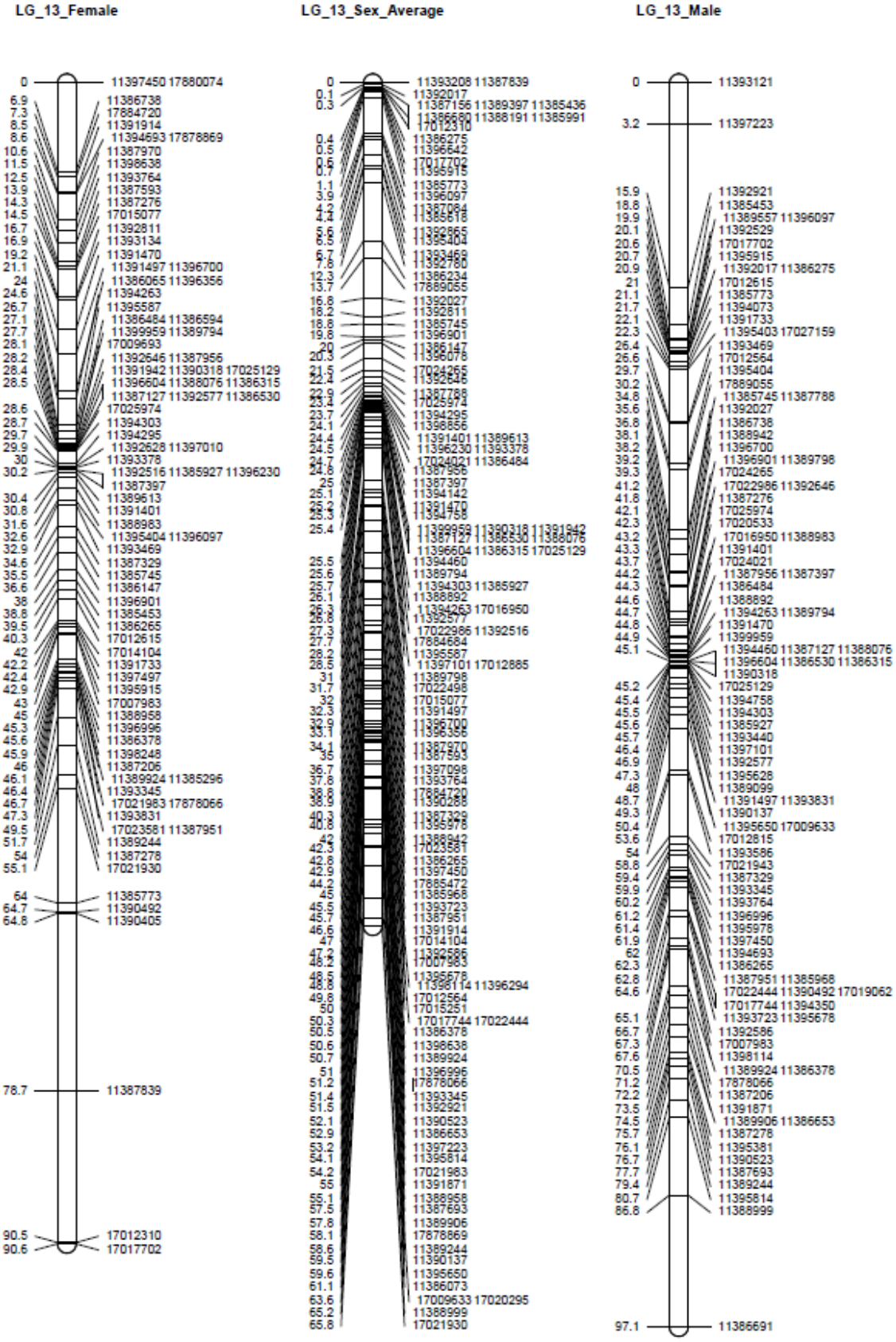

LG_8_Female
0 11392081
0.8 11391980
4.2 11390341
6.9 11390203
7 17010271 11395745 17010025
8.1 11397697
8.4 11397205
8.8 11391994
9 11393156
10.5 11389850
11.7 11394216
15 11387731
16.1 11392009 11396952
16.2 11386848
16.3 11397738
17.3 11387277
17.4 11396935 11395724
20.4 11386379 11391523
20.9 11390071
21.5 17027150 11389958
21.6 11395782 11395565
21.7 11396523
21.8 11393395
22.1 11393951
22.5 17025409 17025207 17018208
23.4 17020714
26.9 11390493
27.5 17882707
28.4 11393123
28.9 11387202
29.3 11392198
29.4 11386402
32.3 11389176
34.7 17017295
37.9 11397499
38 11388225
39.2 11391581 11390260
39.5 17021543 17013273
39.9 11394425
40.1 11386233
40.7 11392344
41.9 11393765
42 11396885
42.4 11397661 17887997 11392285
17017168 11393285 11393294
42.5 17019354 17019346
42.6 11387493 17015386 11393843
42.7 11385878 11395322 17024338
11389899
42.8 11396338 11396470
42.9 11393292 11387289 17016790
11392448 17015188 11396515
11396600 11386640 11393071
17025589 17017289 11393737
11385340 11393779 11397231
11395391
43 11387181 11390284 11396064
43.1 17007214 17021591
43.2 17012365 17013175 17006854
43.4 11397191
44.3 17006204
45 17017027 17878517 11394775
45.8 11392420
45.9 11386936
46.1 17026404
46.6 17012112
46.7 11392018
47.1 11392571
47.8 11387676
48 11392310 11397362
48.2 17021760
49.1 11386712
50.8 11389726 11395777
LG_8_Sex_Average
0 11390341
3.3 11392081
4.3 11388875
5.4 11391980
5.9 11386487
6.6 11395428
6.9 11391523
7 11390243 11397697
8.6 17886994
9 11395294
11.2 11390035
12.6 11387448
13.1 11385701
13.3 11393156
14 11393123
14.3 17020196
14.7 11398971
15.5 17026600
15.8 11386848
16.5 11398501 11385683
17 11395724 11390493
17.3 11395565
17.4 17884589
17.5 11387090
17.8 11390071
18 11398523
18.1 17027150
18.3 11393395
18.5 11389958
18.9 11386379 17025409
19.5 11392009 11396952
19.9 11385776
20.3 11393078
20.7 11387202
21.1 11395782
21.3 17013372
23.1 11393951
24.7 11387731
26.3 11386402 11389850
26.9 11392198
27.8 11392725
27.9 11391994
28.2 17010271
29.1 11395745
29.3 11390203
29.8 17020714
32.1 11394433
33 17017295
33.8 11395777
36.4 11389756
36.7 11398266
37.1 17021760
38.9 11386712
39.4 11396466
39.9 11392344
40.1 11397949
40.2 17012112
40.3 11394425
40.4 11386936
40.8 17013756
40.9 11397362
41 17878517
41.2 11391581
41.4 11386163
41.7 17026922 17882620
41.8 11396338 17007214
42 11393292
42.4 17021591 11396885
42.5 17006854 11393294 11397661
42.6 17019354 17019346 11396064
11394016 17016790 11390284
42.7 17017168 11387181 11387289
42.8 17012365 11387493 11396470
42.9 11393843
43 11395322 17015386 11389899
43.1 11392448 11385878 17024338
17015188
43.2 17013175 11393285 17016874
17025589
43.3 17017289
43.5 11393765
43.7 17008204
43.9 17021730
44.2 11397191
44.3 11396438
44.4 17017027
44.6 11394775
45.6 17026404
45.7 17022372
45.9 11392070
46.3 11395444
46.7 17013273
47.1 11395391
47.6 11387502
48.2 17021543
LG_8_Male
0 11390243
1.8 11388875
2.1 11386487
2.4 11385701 11397697
2.5 11395428
2.8 11392346
3 11391980
3.1 17886994
3.4 11392081
4.5 11395294
6.5 11389176
8.9 11390035
9.6 11390260
10.3 11392571
11.1 11388225
12.9 11393156
13 11398971 11386379
14.3 11394659 17013227
14.5 11395782
14.7 11385776
14.9 11387090
15.1 11397205
15.6 11395724
15.7 11390071
16.2 11393395 11392018
16.5 11392009
16.6 11390259 11390493 11396952
16.8 17026768 17018208 17025207
17025566 17022463 17025409
16.9 11385683 11394216 17013372
17 11397429
17.3 11396935
17.4 11389958
17.5 11386848
17.6 11398501
17.7 17020196
18.5 17020234
21.2 17020714
22 11387277
22.2 11387731
22.9 11389684
23.3 17017295
24 11395745
24.1 11393951
24.7 11392725
25.4 11392420 17010271
26.5 11394433
29.5 17010025
30.7 11392310
30.8 11392198
31 17021760
31.9 11392354
32.8 11386402
33.1 11389756
33.4 11396466
33.8 11396470 11386640 11387493
11397949
34.1 11386936 11387809
34.7 11395264
34.9 17006854
35.5 11385878
35.8 11386163
36.3 17025589
36.9 11395444
37.2 11394519
39.4 17022372
39.8 11387502 11389699 17012112
39.9 11386712 17887997 11391581
11395322 11394016 11393737
17017168 11393071 11387676
11396515 11393765 17021730
11398238 17016790 11387181
11385340 11396600 17015386
11392448 11397231 11394775
40 17019346 11394425
40.2 11392070
40.3 11396885
40.4 11397362 17019354
41.6 11397191
43.4 11389956
44.3 11398266
47.8 11397738
52.8 11395391

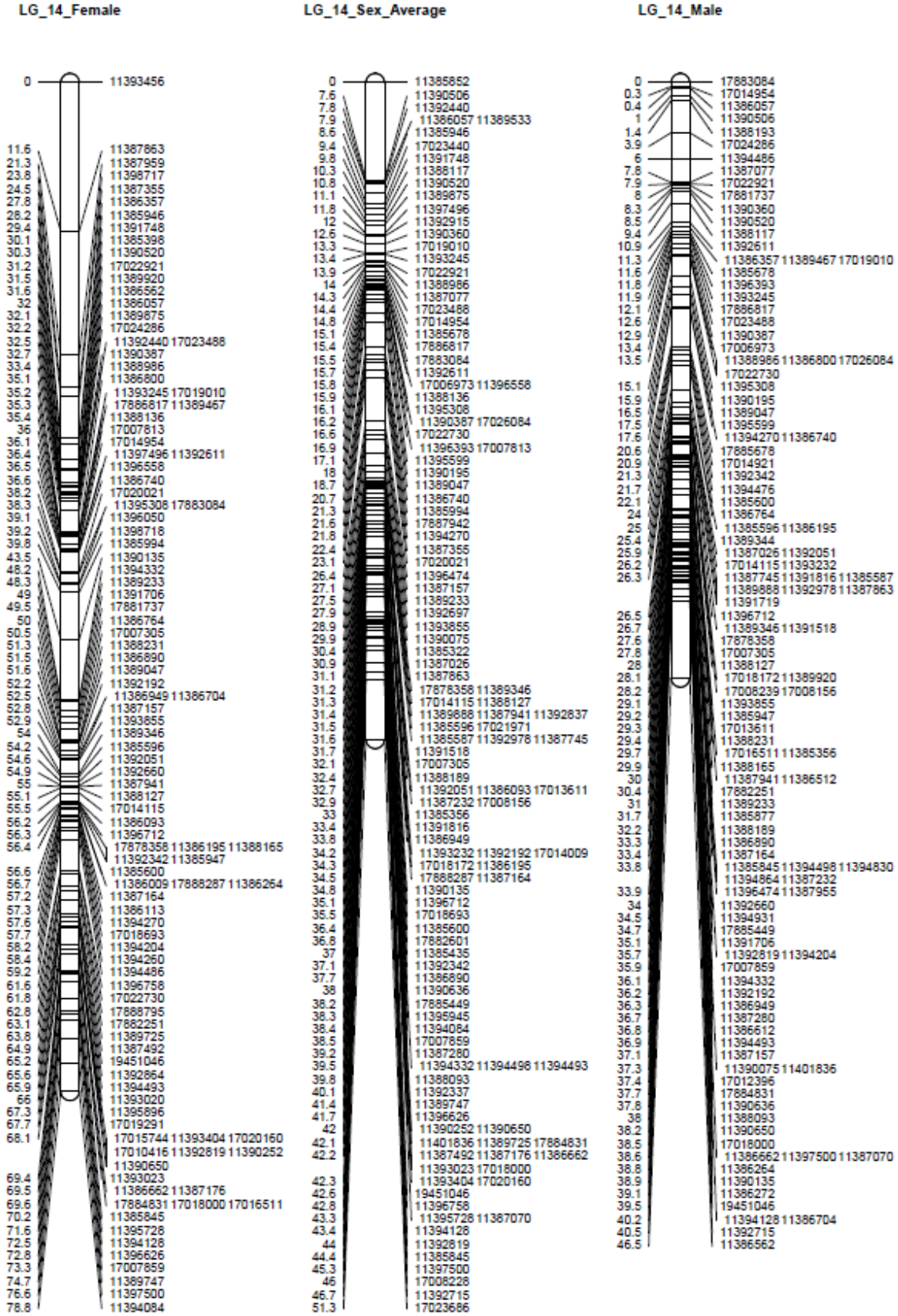

LG_9_Female
0 11392260
2 11387579
6.3 11395530
10.6 11390535
13.7 11396341
20.3 11385328
20.4 11391647
27.6 11396937
29.7 17015656
30 17018981
33.6 17007005
34.4 11390575 11394887
34.7 17017281
39 17886611
40.1 11389219
40.7 11386505
42.1 11386156
43.5 11395488
43.9 11393850 11398517
44.2 17886456
44.4 11385407
44.5 17021564
44.6 11386666 17879613
44.7 17013955 17023654 11395634
11397209 17021489 17883039
17007039 17017507 11385594
17023163
44.8 11388124
45 17017063 11391713
45.1 11387643
45.3 11389603 11389203
45.4 11387453 11391526
45.8 11386453 17015631
45.9 11395921 11394138
46 11385614 11396258
46.1 17013103
46.2 11386331
46.3 17015102 11401503
46.4 11392888
46.5 11396092
46.6 17015012
46.7 11389018
46.8 17018852
46.9 11387856
47.2 11389593
47.3 11393919
47.5 17013693 17013589 17021236
47.6 17023693 11396555
47.8 11386752 11393176
48.9 11393250 11401383 17021830
49 11395840
50.6 11396062
51.3 11389793
51.5 11389209
52.2 11392668
52.7 11392892
53.5 11387023
53.7 11389639
55.8 11398257
56 11395914
56.1 11392172
56.7 11387603
59.8 11391757
66.5 17021144
66.7 17009863
83.2 11387739
84.5 11388153
LG_9_Sex_Average
0 11396341
5.6 17009951
7 17025182 17012616 17889034
17024146
7.9 17022897
8 17010042
8.2 17882557
12 11399468
12.2 17018981
12.6 11387716
15.3 11391647
17.2 11390535
17.9 11387579
18.9 11388153
20.2 11396062
22.6 17013103
23.5 11393176
24 11389018
24.1 17013693 17013589 17023693
24.2 17018852
24.3 17021236
24.5 11387643
25.4 11388124
25.5 11395914
25.9 11395921
26.5 17015631 11386453
26.9 11386505
27.1 11394138
27.3 11385614 17017063
27.8 11391526
27.9 11396157
28 11389203
28.1 11391757
28.2 11396937
28.3 11391713
28.5 17886456
28.6 11393850
28.8 11387603
29.1 11385407
29.2 11397930
29.7 11385620
30.1 11389209
30.5 11393919
30.7 17007262
30.9 11386156
31 11389803
31.3 11389793
32.4 17013955 17015152 17022800
32.9 17883039 17017507 11391756
33 17023101 11386331
33.7 17021666 11401503
34.2 11388874
34.3 17886611
34.5 11389639
35.8 11392668
37.3 11401383
37.8 11390575
38 11398257
38.3 17017281
38.4 11398517
38.7 17021830
39.5 17007005
41.5 11393961
LG_9_Male
0 11389606
5.3 17013103 11389593 11393176
11392888 17018852 17886722
17013589
5.5 17013693
6.1 17009951
6.9 17025182 17012616 17024146
17889034
7.7 17022897
7.8 17010042
8.3 17882557
9.4 11387678
11 17021144 17018981
11.9 11399468
13.2 11387716
15.2 11401383
16.5 11391647
17 17021564
17.9 11390535
19.1 11387579
19.4 11388153
19.9 11386666
21.3 11392260
24.4 11389803
24.5 11395914
24.8 11394887
27.9 11392668 11385328
28.6 11395921
28.8 11386505 17015631 11386453
29.1 11385620 11394138 11393250
11397930
29.2 17021489
29.3 11392892 11387453
29.6 11391526 11391757
29.7 11385614 11389209
30 11390210 11396157
30.1 11387603 11391713
30.2 11386752 11389203 11393850
30.3 17017063
31.1 11390575
31.2 17023654
31.6 11386331
32.2 17023163
33 11389793 11394577
33.5 17883039 17017507
33.7 11391756
34.7 11396092
35 11398517
35.1 11396258
35.4 17015012
36.2 11385407
36.5 11401503
36.6 17021666 11389639
37.2 11388874
37.3 17007005
38.3 11392172
39.5 11395840
47.7 17015656
48.5 11387549
56.2 17015102
60.6 11396062

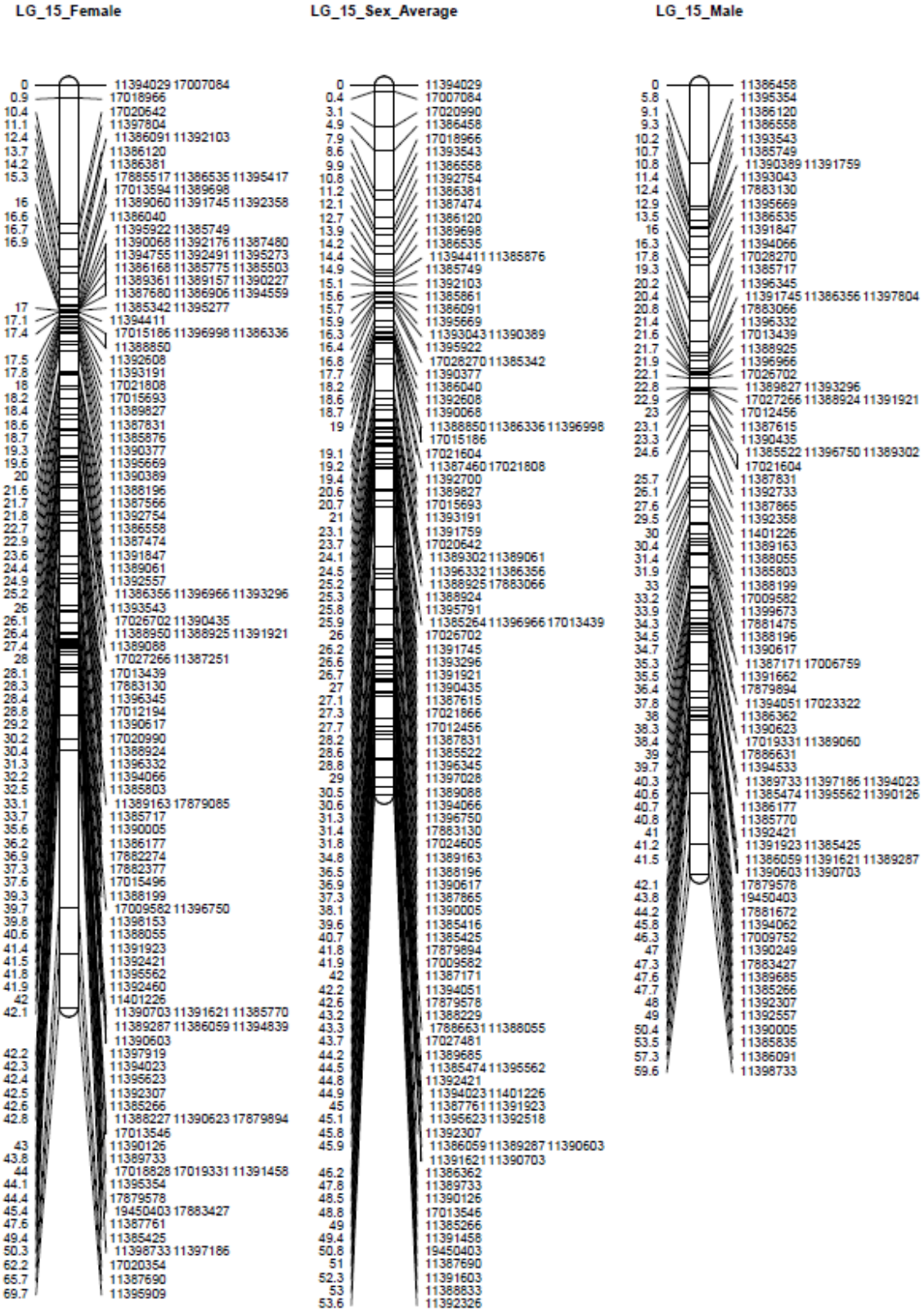

LG_10_Female
LG_10_Sex_Average
LG_10_Male
0
0
0
0
0
17018752
14.1
11386801
33.6
34.2
40.3
42
45.2
45.6
46.2
47.6
50
50.6
51.1
52.2
52.3
52.5
52.7
52.8
11388046
11397550
11397515
11392465
11396029
11389985
11395436
11397706
11385771 11387575
11388029
11387247
17006013
17885656
11393937
11393282
17014962 17020788 11385568
11395935 11385827 17025283
17024597 17019070 11393008
11396244 11392968 11389319
11397933 17023115 11391394
11386355 17023646 17017119
11396755 17020169 17007817
11392224
54.8
55
55.2
57.3
57.6
58
58.4
59.5
59.7
60.4
61.4
62.1
62.3
62.4
11393743
11387578
11395680
11388969
11387151
11392106
11397780 11386488
11395378
17878308 11398241
11394406
11388879
11387882
11391677
17019174 17023362 17021681
17889323
70.9
71.3
11391904
17025027
81.1
81.5
11391502
11395149
103.3
11387085

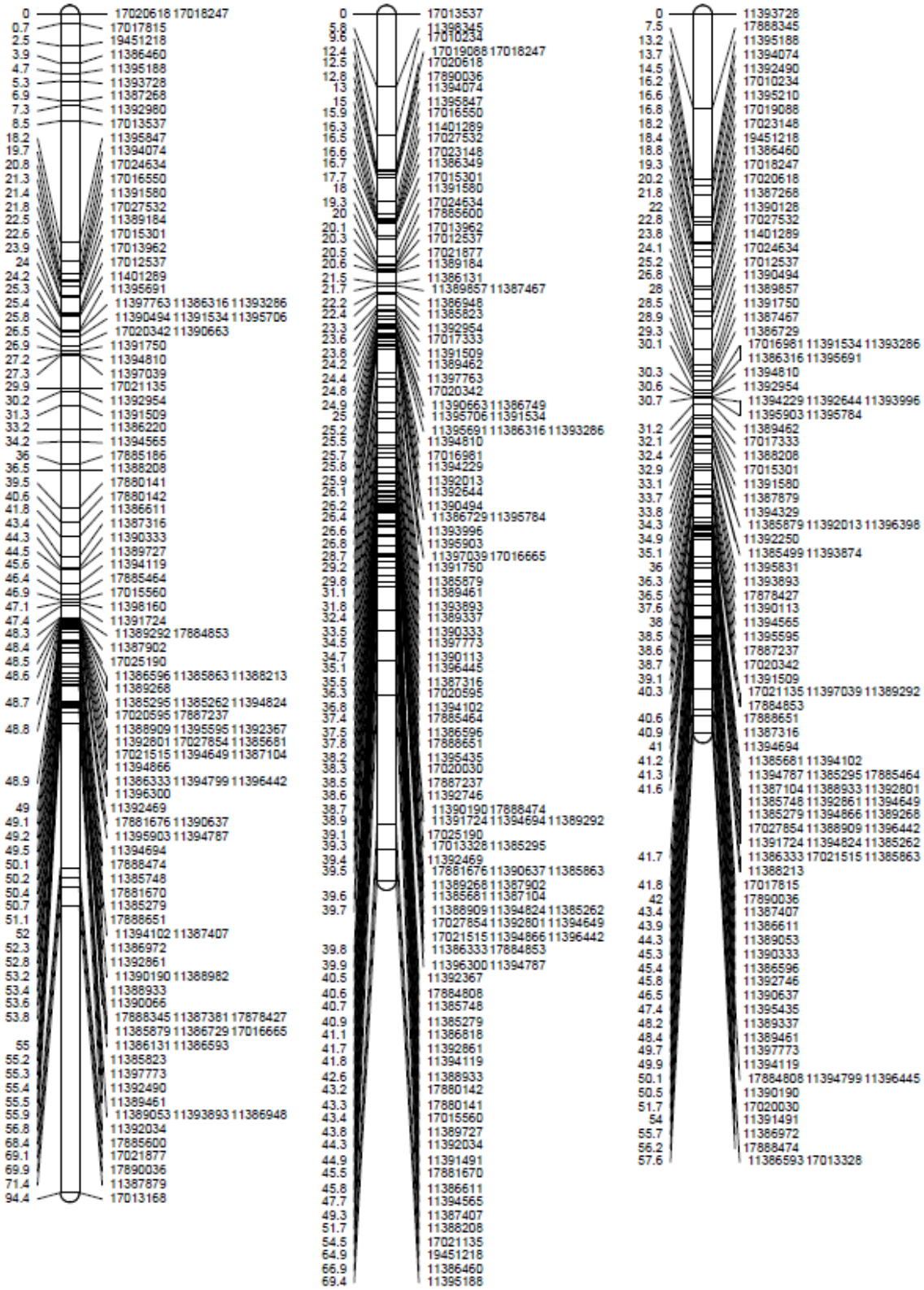

LG_11_Female
LG_11_Sex_Average
LG_11_Male

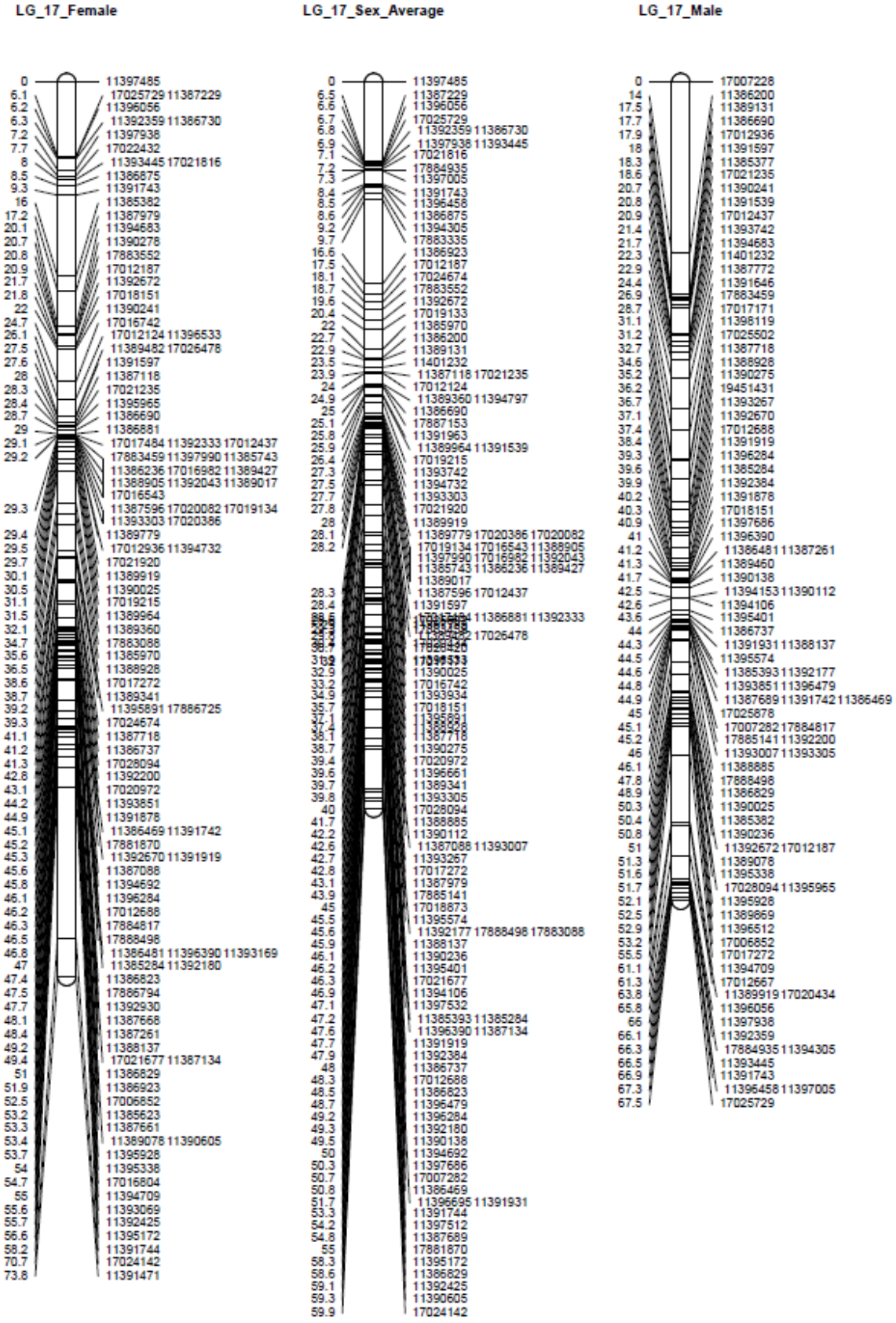

LG_12_Female
LG_12_Sex_Average
LG_12_Male

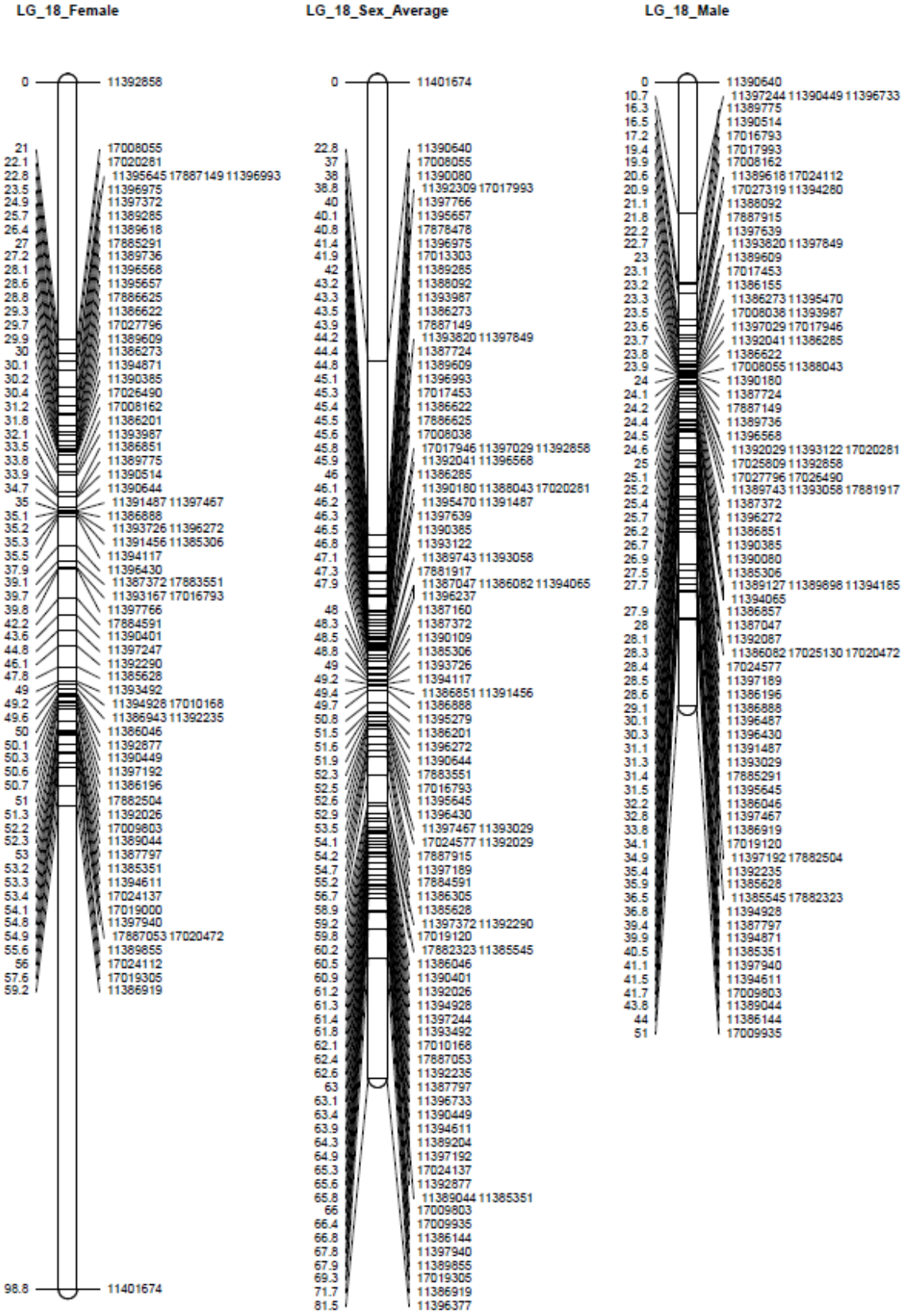

LG_13_Female
LG_13_Sex_Average
LG_13_Male

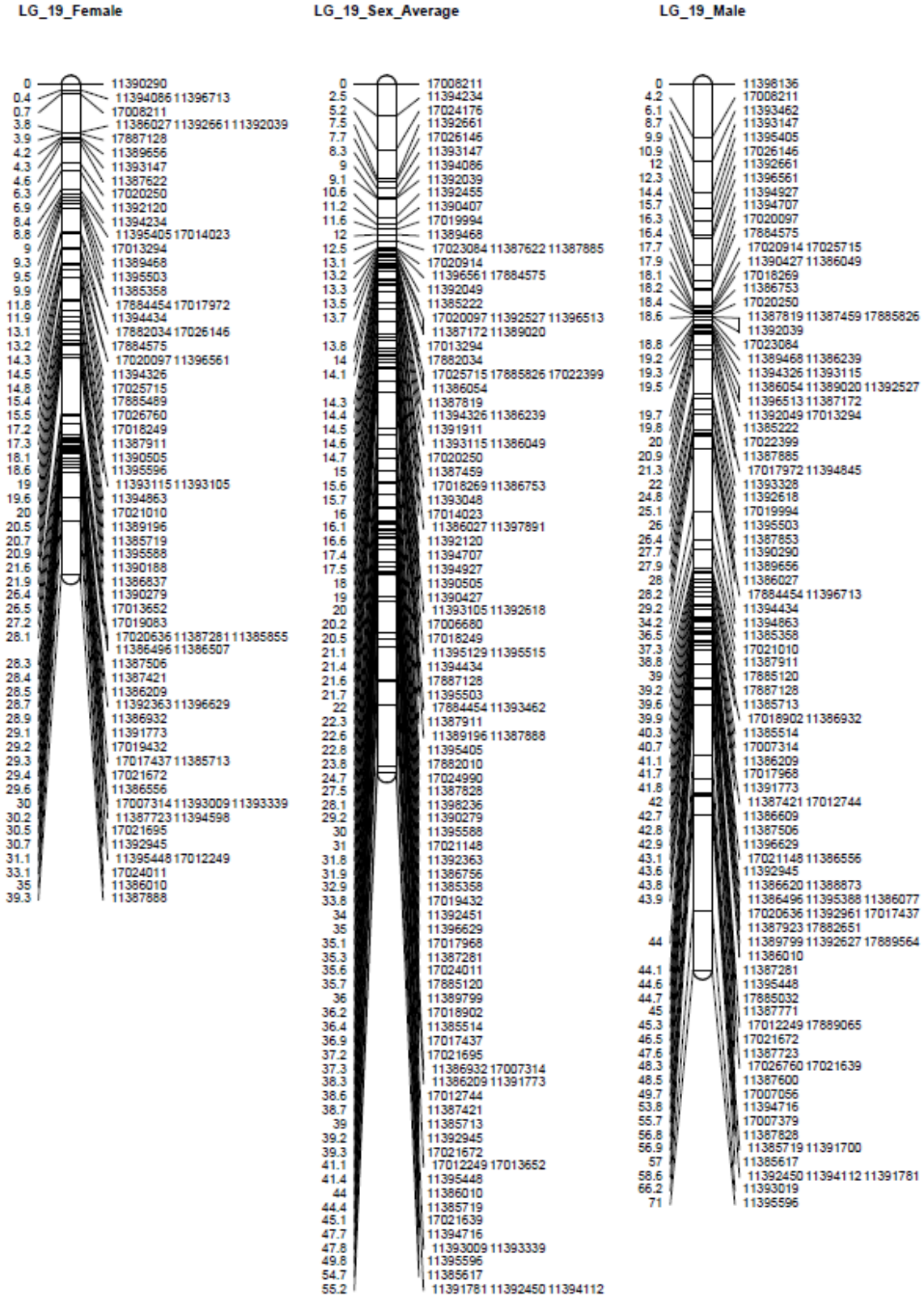

LG_14_Female
LG_14_Sex_Average
LG_14_Male

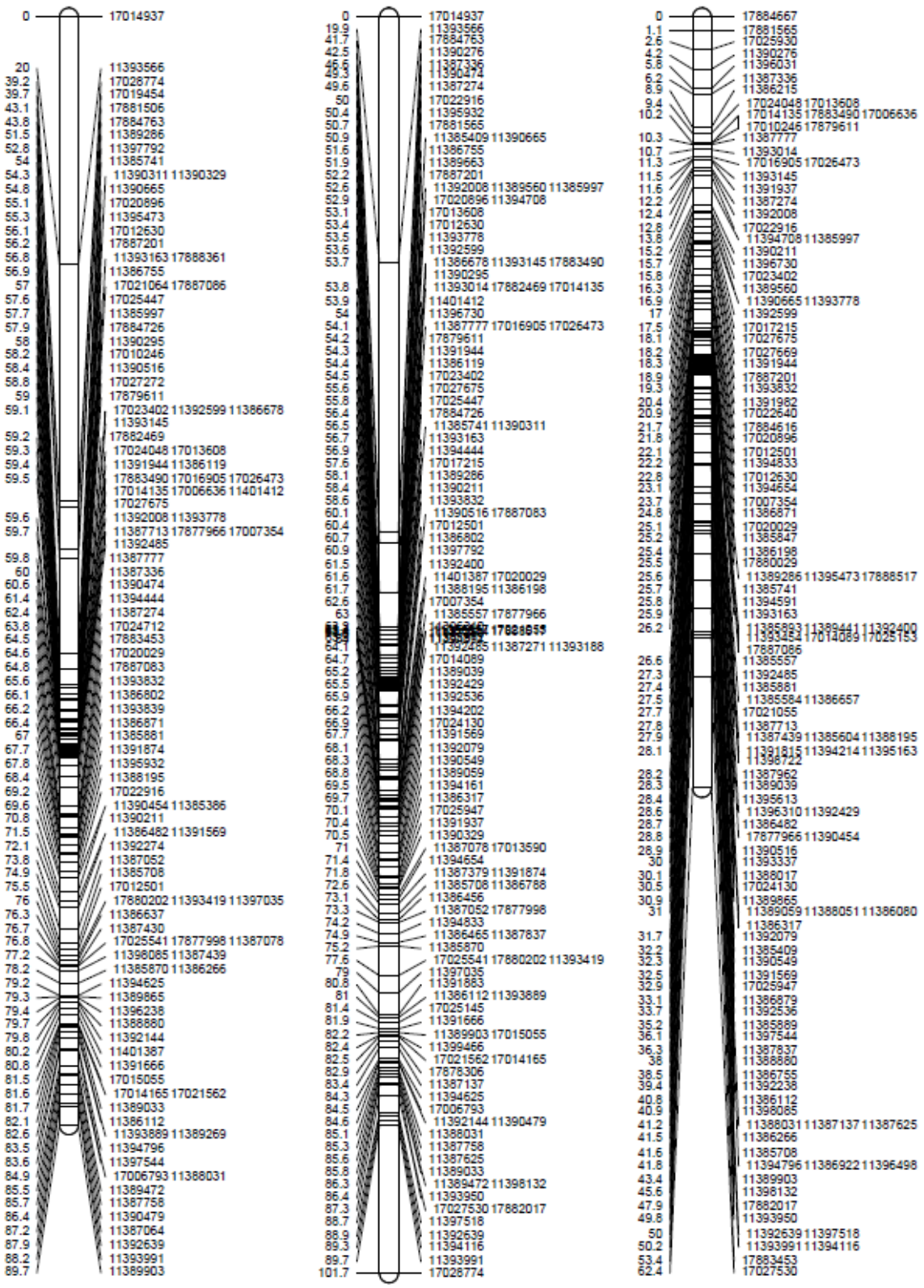

LG_15_Female
LG_15_Sex_Average
LG_15_Male

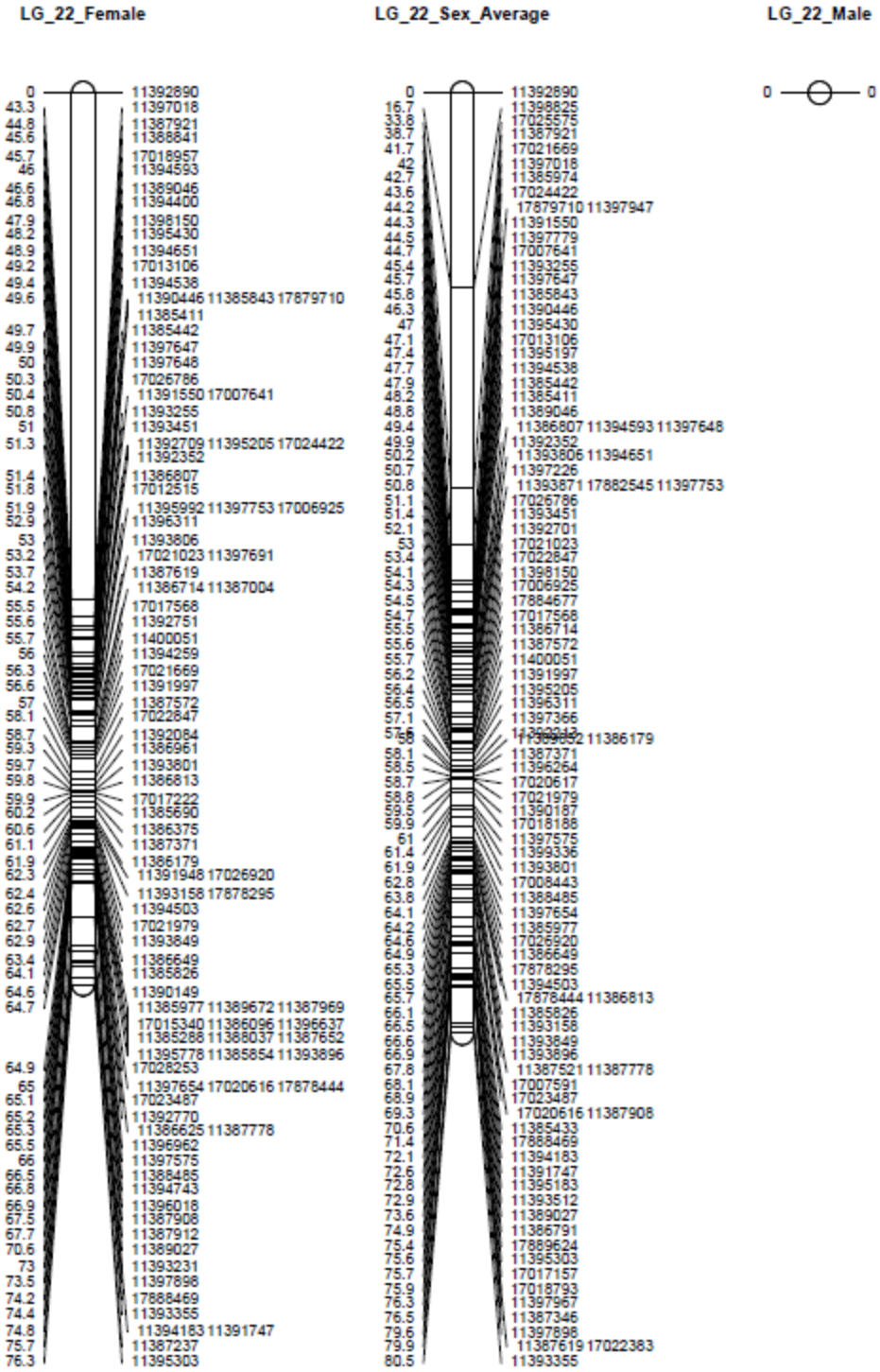

LG_16_Female
LG_16_Sex_Average
LG_16_Male

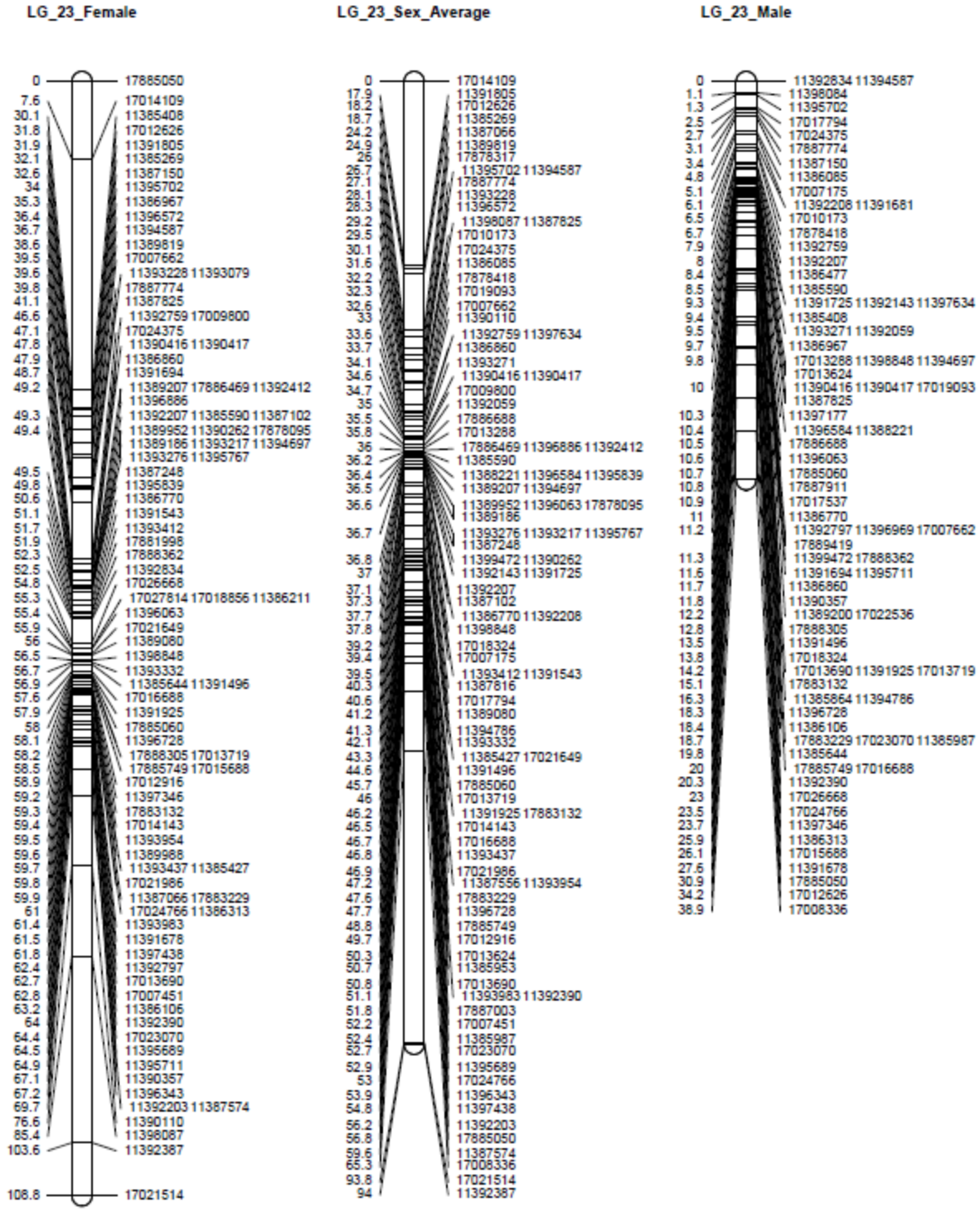

LG_17_Female
LG_17_Sex_Average
LG_17_Male


LG_18_Female
0 11392858
21 17008055
22.1 17020281
22.8 11395645 17887149 11396993
23.5 11396975
24.9 11397372
25.7 11389285
26.4 11389618
27 17885291
27.2 11389736
28.1 11396568
28.6 11395657
28.8 17886625
29.3 11386622
29.7 17027796
29.9 11389609
30 11386273
30.1 11394871
30.2 11390385
30.4 17026490
31.2 17008162
31.8 11386201
32.1 11393987
33.5 11386851
33.8 11389775
33.9 11390514
34.7 11390644
35 11391487 11397467
35.1 11386888
35.2 11393726 11396272
35.3 11391456 11385306
35.5 11394117
37.9 11396430
39.1 11387372 17883551
39.7 11393167 17016793
39.8 11397766
42.2 17884591
43.6 11390401
44.8 11397247
46.1 11392290
47.8 11385628
49 11393492
49.2 11394928 17010168
49.6 11386943 11392235
50 11386046
50.1 11392877
50.3 11390449
50.6 11397192
50.7 11386196
51 17882504
51.3 11392026
52.2 17009803
52.3 11389044
53 11387797
53.2 11385351
53.3 11394611
53.4 17024137
54.1 17019000
54.8 11397940
54.9 17887053 17020472
55.6 11389855
56 17024112
57.6 17019305
59.2 11386919
98.8 11401674
LG_18_Sex_Average
0 11401674
22.8 11390640
37 17008055
38 11390080
38.8 11392309 17017993
40 11397766
40.1 11395657
40.8 17878478
41.4 11396975
41.9 17013303
42 11389285
43.2 11388092
43.3 11393987
43.5 11386273
43.9 17887149
44.2 11393820 11397849
44.4 11387724
44.8 11389609
45.1 11396993
45.3 17017453
45.4 11386622
45.5 17886625
45.6 17008038
45.8 17017946 11397029 11392858
45.9 11392041 11396568
46 11386285
46.1 11390180 11388043 17020281
46.2 11395470 11391487
46.3 11397639
46.5 11390385
46.8 11393122
47.1 11389743 11393058
47.3 17881917
47.9 11387047 11386082 11394065
11396237
48 11387160
48.3 11387372
48.5 11390109
48.8 11385306
49 11393726
49.2 11394117
49.4 11386851 11391456
49.7 11386888
50.8 11395279
51.5 11386201
51.6 11396272
51.9 11390644
52.3 17883551
52.5 17016793
52.6 11395645
52.9 11396430
53.5 11397467 11393029
54.1 17024577 11392029
54.2 17887915
54.7 11397189
55.2 17884591
56.7 11386305
58.9 11385628
59.2 11397372 11392290
59.8 17019120
60.2 17882323 11385545
60.5 11386046
60.9 11390401
61.2 11392026
61.3 11394928
61.4 11397244
61.8 11393492
62.1 17010168
62.4 17887053
62.6 11392235
63 11387797
63.1 11396733
63.4 11390449
63.9 11394611
64.3 11389204
64.9 11397192
65.3 17024137
65.6 11392877
65.8 11389044 11385351
66 17009803
66.4 17009935
66.8 11386144
67.8 11397940
67.9 11389855
69.3 17019305
71.7 11386919
81.5 11396377
LG_18_Male
0 11390640
10.7 11397244 11390449 11396733
16.3 11389775
16.5 11390514
17.2 17016793
19.4 17017993
19.9 17008162
20.6 11389618 17024112
20.9 17027319 11394280
21.1 11388092
21.8 17887915
22.2 11397639
22.7 11393820 11397849
23 11389609
23.1 17017453
23.2 11386155
23.3 11386273 11395470
23.5 17008038 11393987
23.6 11397029 17017946
23.7 11392041 11386285
23.8 11386622
23.9 17008055 11388043
24 11390180
24.1 11387724
24.2 17887149
24.4 11389736
24.5 11396568
24.6 11392029 11393122 17020281
25 17025809 11392858
25.1 17027796 17026490
25.2 11389743 11393058 17881917
25.4 11387372
25.7 11396272
26.2 11386851
26.7 11390385
26.9 11390080
27.5 11385306
27.7 11389127 11389898 11394185
11394065
27.9 11386857
28 11387047
28.1 11392087
28.3 11386082 17025130 17020472
28.4 17024577
28.5 11397189
28.6 11386196
29.1 11386888
30.1 11396487
30.3 11396430
31.1 11391487
31.3 11393029
31.4 17885291
31.5 11395645
32.2 11386046
32.8 11397467
33.8 11386919
34.1 17019120
34.9 11397192 17882504
35.4 11392235
35.9 11385628
36.5 11385545 17882323
36.8 11394928
39.4 11387797
39.9 11394871
40.5 11385351
41.1 11397940
41.5 11394611
41.7 17009803
43.8 11389044
44 11386144
51 17009935

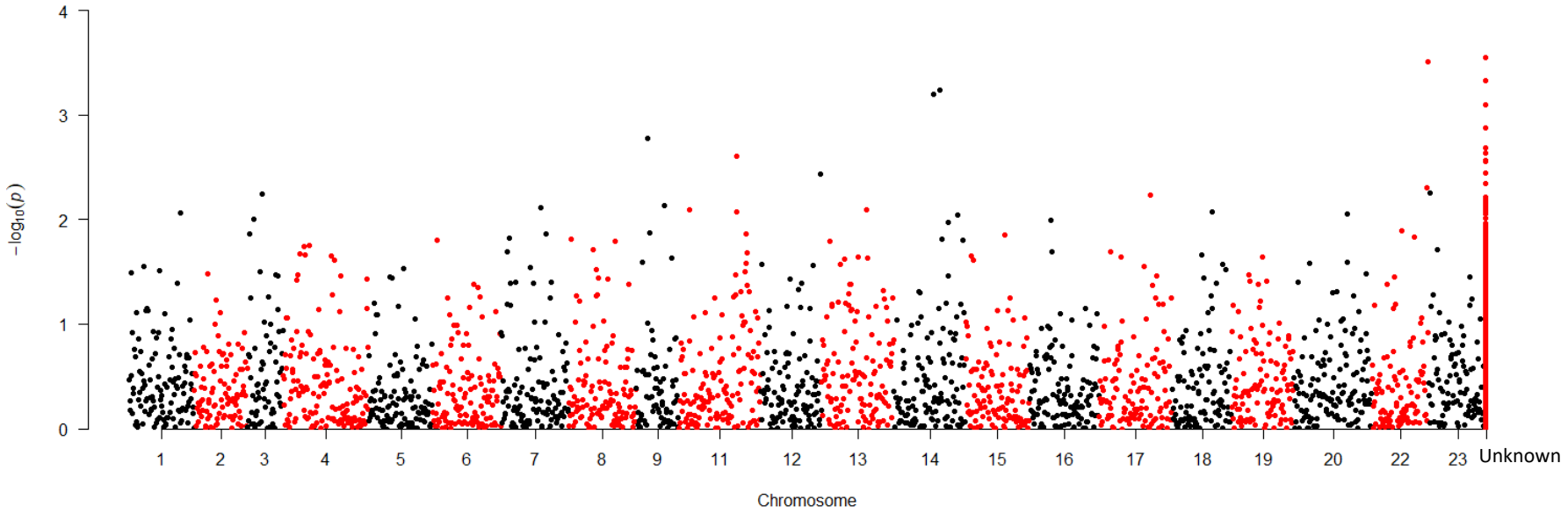

LG_19_Female
0 11390290
0.4 11394086 11396713
0.7 17008211
3.8 11386027 11392661 11392039
3.9 17887128
4.2 11389656
4.3 11393147
4.6 11387622
6.3 17020250
6.9 11392120
8.4 11394234
8.8 11395405 17014023
9 17013294
9.3 11389468
9.5 11395503
9.9 11385358
11.8 17884454 17017972
11.9 11394434
13.1 17882034 17026146
13.2 17884575
14.3 17020097 11396561
14.5 11394326
14.8 17025715
15.4 17885489
15.5 17026760
17.2 17018249
17.3 11387911
18.1 11390505
18.6 11395596
19 11393115 11393105
19.6 11394863
20 17021010
20.5 11389196
20.7 11385719
20.9 11395588
21.6 11390188
21.9 11386837
26.4 11390279
26.5 17013652
27.2 17019083
28.1 17020536 11387281 11385855 11386496 11386507
28.3 11387506
28.4 11387421
28.5 11386209
28.7 11392363 11396629
28.9 11386932
29.1 11391773
29.2 17019432
29.3 17017437 11385713
29.4 17021672
29.6 11386556
30 17007314 11393009 11393339
30.2 11387723 11394598
30.5 17021695
30.7 11392945
31.1 11395448 17012249
33.1 17024011
35 11386010
39.3 11387888
LG_19_Sex_Average
0 17008211
2.5 11394234
5.2 17024176
7.5 11392661
7.7 17026146
8.3 11393147
9 11394086
9.1 11392039
10.6 11392455
11.2 11390407
11.6 17019994
12 11389468
12.5 17023084 11387622 11387885
13.1 17020914
13.2 11396561 17884575
13.3 11392049
13.5 11385222
13.7 17020097 11392527 11396513 11387172 11389020
13.8 17013294
14 17882034
14.1 17025715 17885826 17022399 11386054
14.3 11387819
14.4 11394326 11386239
14.5 11391911
14.6 11393115 11386049
14.7 17020250
15 11387459
15.6 17018269 11386753
15.7 11393048
16 17014023
16.1 11386027 11397891
16.6 11392120
17.4 11394707
17.5 11394927
18 11390505
19 11390427
20 11393105 11392618
20.2 17006660
20.5 17018249
21.1 11395129 11395515
21.4 11394434
21.6 17887128
21.7 11395503
22 17884454 11393462
22.3 11387911
22.6 11389196 11387888
22.8 11395405
23.8 17882010
24.7 17024990
27.5 11387828
28.1 11398236
29.2 11390279
30 11395588
31 17021148
31.8 11392363
31.9 11386756
32.9 11385358
33.8 17019432
34 11392451
35 11396629
35.1 17017968
35.3 11387281
35.6 17024011
35.7 17885120
36 11389799
36.2 17018902
36.4 11385514
36.9 17017437
37.2 17021695
37.3 11386932 17007314
38.3 11386209 11391773
38.6 17012744
38.7 11387421
39 11385713
39.2 11392945
39.3 17021672
41.1 17012249 17013652
41.4 11395448
44 11386010
44.4 11385719
45.1 17021639
47.7 11394716
47.8 11393009 11393339
49.8 11395596
54.7 11385617
55.2 11391781 11392450 11394112
LG_19_Male
0 11398136
4.2 17008211
6.1 11393462
8.7 11393147
9.9 11395405
10.9 17026146
12 11392661
12.3 11396561
14.4 11394927
15.7 11394707
16.3 17020097
16.4 17884575
17.7 17020914 17025715
17.9 11390427 11386049
18.1 17018269
18.2 11386753
18.4 17020250
18.6 11387819 11387459 17885826 11392039
18.8 17023084
19.2 11389468 11386239
19.3 11394326 11393115
19.5 11386054 11389020 11392527 11396513 11387172
19.7 11392049 17013294
19.8 11385222
20 17022399
20.9 11387885
21.3 17017972 11394845
22 11393328
24.8 11392618
25.1 17019994
26 11395503
26.4 11387853
27.7 11390290
27.9 11389656
28 11386027
28.2 17884454 11396713
29.2 11394434
34.2 11394863
36.5 11385358
37.3 17021010
38.8 11387911
39 17885120
39.2 17887128
39.6 11385713
39.9 17018902 11386932
40.3 11385514
40.7 17007314
41.1 11386209
41.7 17017968
41.8 11391773
42 11387421 17012744
42.7 11386609
42.8 11387506
42.9 11396629
43.1 17021148 11386556
43.6 11392945
43.8 11386620 11388873
43.9 11386496 11395388 11386077 17020636 11392961 17017437 11387923 17882651
44 11389799 11392627 17889564 11386010
44.1 11387281
44.6 11395448
44.7 17885032
45 11387771
45.3 17012249 17889065
46.5 17021672
47.6 11387723
48.3 17026760 17021639
48.5 11387600
49.7 17007056
53.8 11394716
55.7 17007379
56.8 11387828
56.9 11385719 11391700
57 11385617
58.6 11392450 11394112 11391781
66.2 11393019
71 11395596


LG_20_Female
LG_20_Sex_Average
LG_20_Male


LG_22_Female
LG_22_Sex_Average
LG_22_Male
0 11392890
43.3 11397018
44.8 11387921
45.6 11388841
45.7 17018957
46 11394593
46.6 11389046
46.8 11394400
47.9 11398150
48.2 11395430
48.9 11394651
49.2 17013106
49.4 11394538
49.6 11390446 11385843 17879710 11385411
49.7 11385442
49.9 11397647
50 11397648
50.3 17026786
50.4 11391550 17007641
50.8 11393255
51 11393451
51.3 11392709 11395205 17024422 11392352
51.4 11386807
51.8 17012515
51.9 11395992 11397753 17006925
52.9 11396311
53 11393806
53.2 17021023 11397691
53.7 11387619
54.2 11386714 11387004
55.5 17017568
55.6 11392751
55.7 11400051
56 11394259
56.3 17021669
56.6 11391997
57 11387572
58.1 17022847
58.7 11392084
59.3 11386961
59.7 11393801
59.8 11386813
59.9 17017222
60.2 11385690
60.6 11386375
61.1 11387371
61.9 11386179
62.3 11391948 17026920
62.4 11393158 17878295
62.6 11394503
62.7 17021979
62.9 11393849
63.4 11386649
64.1 11385826
64.6 11390149
64.7 11385977 11389672 11387969 17015340 11386096 11396637 11385288 11388037 11387652 11395778 11385854 11393896
64.9 17028253
65 11397654 17020616 17878444
65.1 17023487
65.2 11392770
65.3 11386625 11387778
65.5 11396962
66 11397575
66.5 11388485
66.8 11394743
66.9 11396018
67.5 11387908
67.7 11387912
70.6 11389027
73 11393231
73.5 11397898
74.2 17888469
74.4 11393355
74.8 11394183 11391747
75.7 11387237
76.3 11395303
0 11392890
16.7 11398825
33.6 17025575
38.7 11387921
41.7 17021669
42 11397018
42.7 11385974
43.6 17024422
44.2 17879710 11397947
44.3 11391550
44.5 11397779
44.7 17007641
45.4 11393255
45.7 11397647
45.8 11385843
46.3 11390446
47 11395430
47.1 17013106
47.4 11395197
47.7 11394538
47.9 11385442
48.2 11385411
48.8 11389046
49.4 11386807 11394593 11397648
49.9 11392352
50.2 11393806 11394651
50.7 11397226
50.8 11393871 17882545 11397753
51.1 17026786
51.4 11393451
52.1 11392701
53 17021023
53.4 17022847
54.1 11398150
54.3 17006925
54.5 17884677
54.7 17017568
55.5 11386714
55.6 11387572
55.7 11400051
56.2 11391997
56.4 11395205
56.5 11396311
57.1 11397366
58 11386179
58.1 11387371
58.5 11396264
58.7 17020617
58.8 17021979
59.5 11390187
59.9 17018188
61 11397575
61.4 11399336
61.9 11393801
62.8 17008443
63.8 11388485
64.1 11397654
64.2 11385977
64.6 17026920
64.9 11386649
65.3 17878295
65.5 11394503
65.7 17878444 11386813
66.1 11385826
66.5 11393158
66.6 11393849
66.9 11393896
67.8 11387521 11387778
68.1 17007591
68.9 17023487
69.3 17020616 11387908
70.6 11385433
71.4 17888469
72.1 11394183
72.6 11391747
72.8 11395183
72.9 11393512
73.6 11389027
74.9 11386791
75.4 17889624
75.6 11395303
75.7 17017157
75.9 17018793
76.3 11397967
76.5 11387346
79.6 11397898
79.9 11387619 17022383
80.5 11393355
0 0

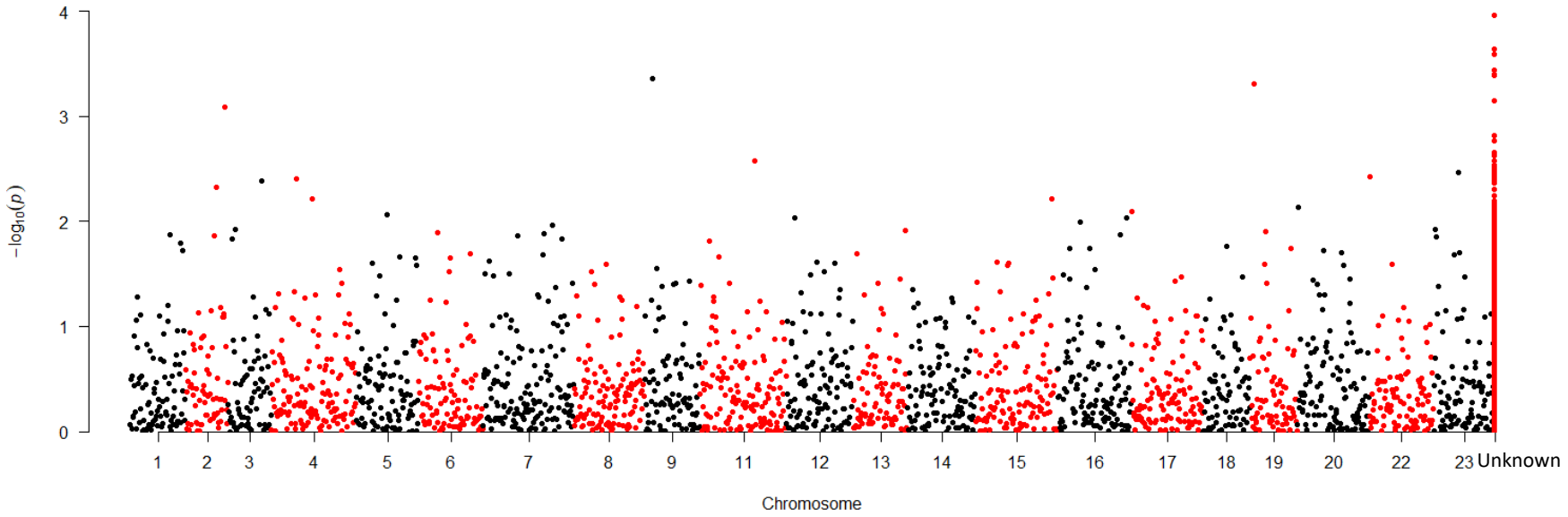

LG_23_Female
0 17885050
7.6 17014109
30.1 11385408
31.8 17012626
31.9 11391805
32.1 11385269
32.6 11387150
34 11395702
35.3 11386967
36.4 11396572
36.7 11394587
38.6 11389819
39.5 17007662
39.6 11393228 11393079
39.8 17887774
41.1 11387825
46.6 11392759 17009800
47.1 17024375
47.8 11390416 11390417
47.9 11386860
48.7 11391694
49.2 11389207 17886469 11392412
11396886
49.3 11392207 11385590 11387102
49.4 11389952 11390262 17878095
11389186 11393217 11394697
11393276 11395767
49.5 11387248
49.8 11395839
50.6 11386770
51.1 11391543
51.7 11393412
51.9 17881998
52.3 17888362
52.5 11392834
54.8 17026668
55.3 17027814 17018856 11386211
55.4 11396063
55.9 17021649
56 11389080
56.5 11398848
56.7 11393332
56.9 11385644 11391496
57.6 17016688
57.9 11391925
58 17885060
58.1 11396728
58.2 17888305 17013719
58.5 17885749 17015688
58.9 17012916
59.2 11397346
59.3 17883132
59.4 17014143
59.5 11393954
59.6 11389988
59.7 11393437 11385427
59.8 17021986
59.9 11387066 17883229
61 17024766 11386313
61.4 11393963
61.5 11391678
61.8 11397438
62.4 11392797
62.7 17013690
62.8 17007451
63.2 11386106
64 11392390
64.4 17023070
64.5 11395689
64.9 11395711
67.1 11390357
67.2 11396343
69.7 11392203 11387574
76.6 11390110
85.4 11398087
103.6 11392387
108.8 17021514
LG_23_Sex_Average
0 17014109
17.9 11391805
18.2 17012626
18.7 11385269
24.2 11387066
24.9 11389819
26 17878317
26.7 11395702 11394587
27.1 17887774
28.1 11393228
28.3 11396572
29.2 11398087 11387825
29.5 17010173
30.1 17024375
31.6 11386085
32.2 17878418
32.3 17019093
32.6 17007662
33 11390110
33.6 11392759 11397634
33.7 11386860
34.1 11393271
34.6 11390416 11390417
34.7 17009800
35 11392059
35.5 17886688
35.8 17013288
36 17886469 11396886 11392412
36.2 11385590
36.4 11388221 11396584 11395839
36.5 11389207 11394697
36.6 11389952 11396063 17878095
11389186
36.7 11393276 11393217 11395767
11387248
36.8 11399472 11390262
37 11392143 11391725
37.1 11392207
37.3 11387102
37.7 11386770 11392208
37.8 11398848
39.2 17018324
39.4 17007175
39.5 11393412 11391543
40.3 11387816
40.6 17017794
41.2 11389080
41.3 11394786
42.1 11393332
43.3 11385427 17021649
44.6 11391496
45.7 17885060
46 17013719
46.2 11391925 17883132
46.5 17014143
46.7 17016688
46.8 11393437
46.9 17021986
47.2 11387556 11393954
47.6 17883229
47.7 11396728
48.8 17885749
49.7 17012916
50.3 17013624
50.7 11385953
50.8 17013690
51.1 11393963 11392390
51.8 17887003
52.2 17007451
52.4 11385987
52.7 17023070
52.9 11395689
53 17024766
53.9 11396343
54.8 11397438
56.2 11392203
56.8 17885050
59.6 11387574
65.3 17008336
93.8 17021514
94 11392387
LG_23_Male
0 11392834 11394587
1.1 11398084
1.3 11395702
2.5 17017794
2.7 17024375
3.1 17887774
3.4 11387150
4.8 11386085
5.1 17007175
6.1 11392208 11391681
6.5 17010173
6.7 17878418
7.9 11392759
8 11392207
8.4 11386477
8.5 11385590
9.3 11391725 11392143 11397634
9.4 11385408
9.5 11393271 11392059
9.7 11386967
9.8 17013288 11398848 11394697
17013624
10 11390416 11390417 17019093
11387825
10.3 11397177
10.4 11396584 11388221
10.5 17886688
10.6 11396063
10.7 17885060
10.8 17887911
10.9 17017537
11 11386770
11.2 11392797 11396969 17007662
17889419
11.3 11399472 17888362
11.6 11391694 11395711
11.7 11386860
11.8 11390357
12.2 11389200 17022536
12.8 17888305
13.5 11391496
13.8 17018324
14.2 17013690 11391925 17013719
15.1 17883132
16.3 11385864 11394786
18.3 11396728
18.4 11386106
18.7 17883229 17023070 11385987
19.8 11385644
20 17885749 17016688
20.3 11392390
23 17026668
23.5 17024766
23.7 11397346
25.9 11386313
26.1 17015688
27.6 11391678
30.9 17885050
34.2 17012626
38.9 17008336

APPENDIX 7



Appendix 7. Manhattan plot based on the sex average linkage map for markers associated with weight. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line.

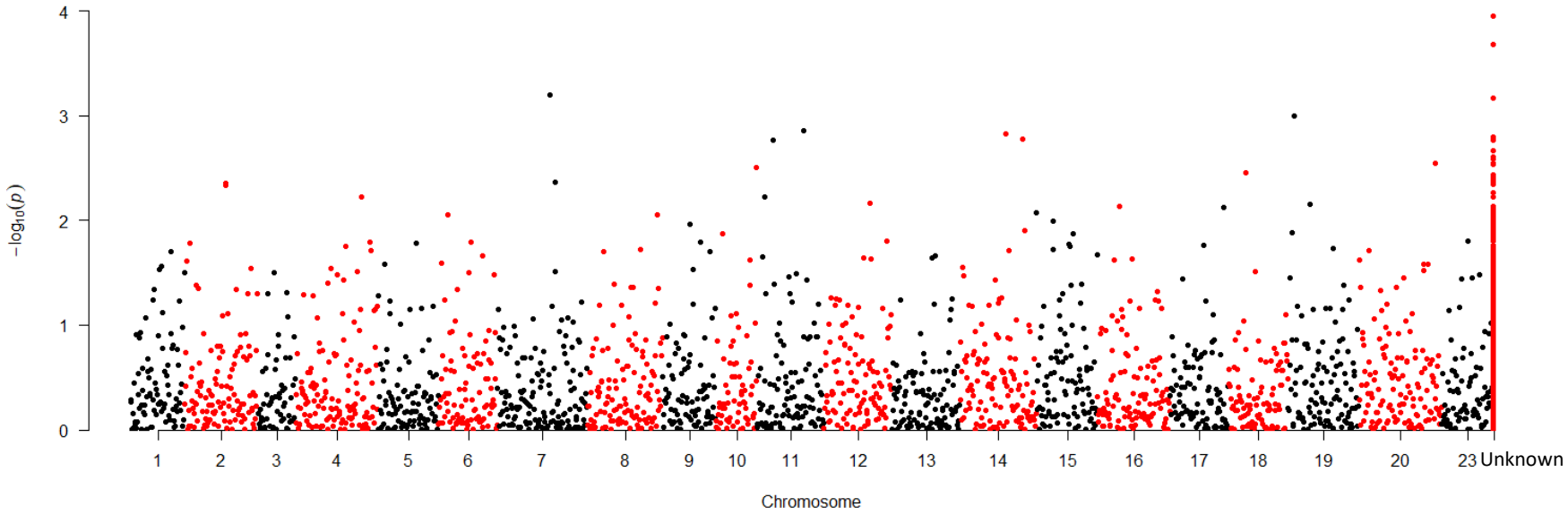
APPENDIX 8



Appendix 8. Manhattan plot based on the female linkage map for markers associated with weight. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line.

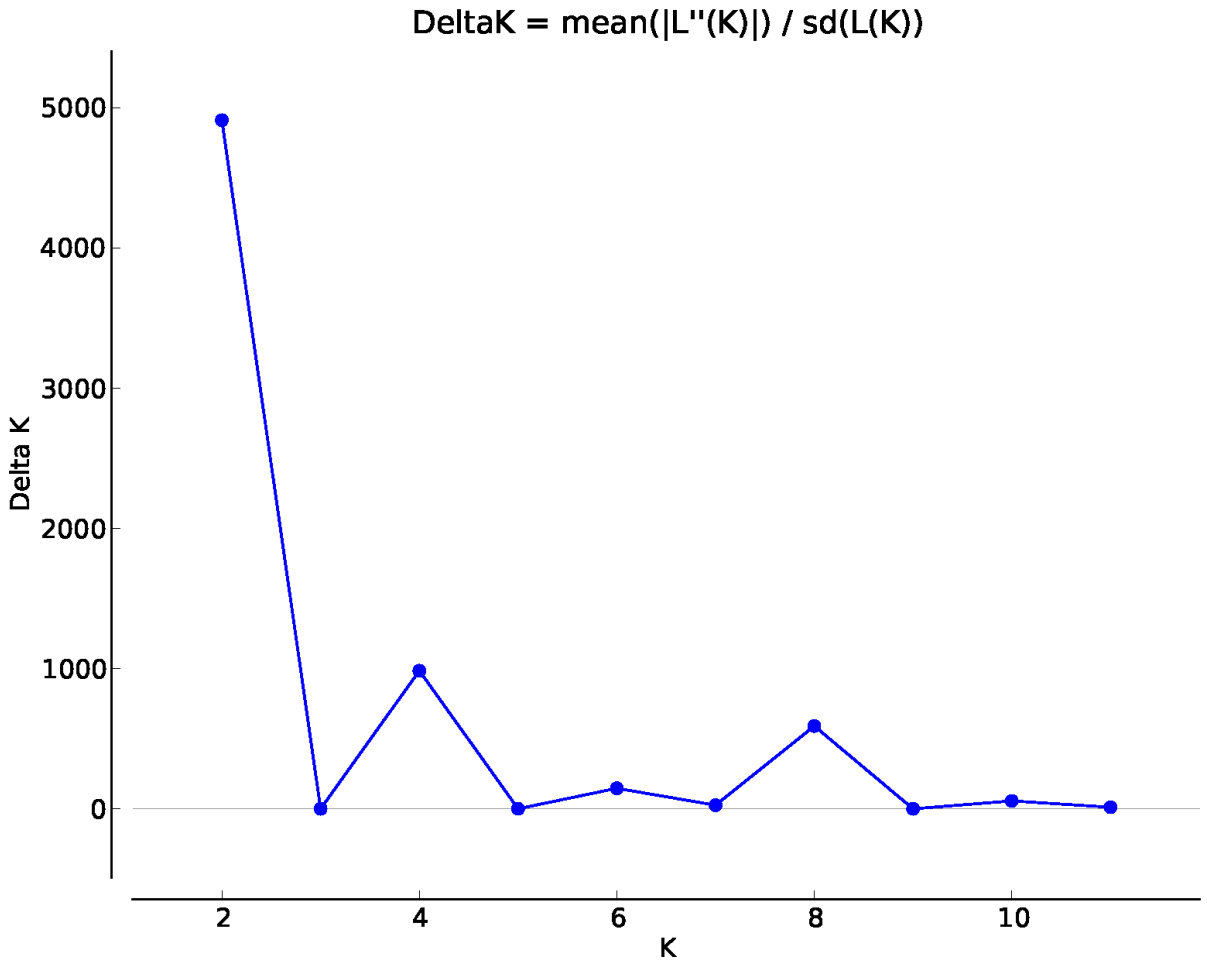
APPENDIX 9

.



Appendix 9. Manhattan plot based on the male linkage map for markers associated with weight. The distribution of p-values for all informative SNPs where the threshold for genome-wide significance at a p-value of $1x10^{-5}$ is denoted by a solid blue line and $1x10^{-10}$ is denoted by a solid red line.

APPENDIX 10



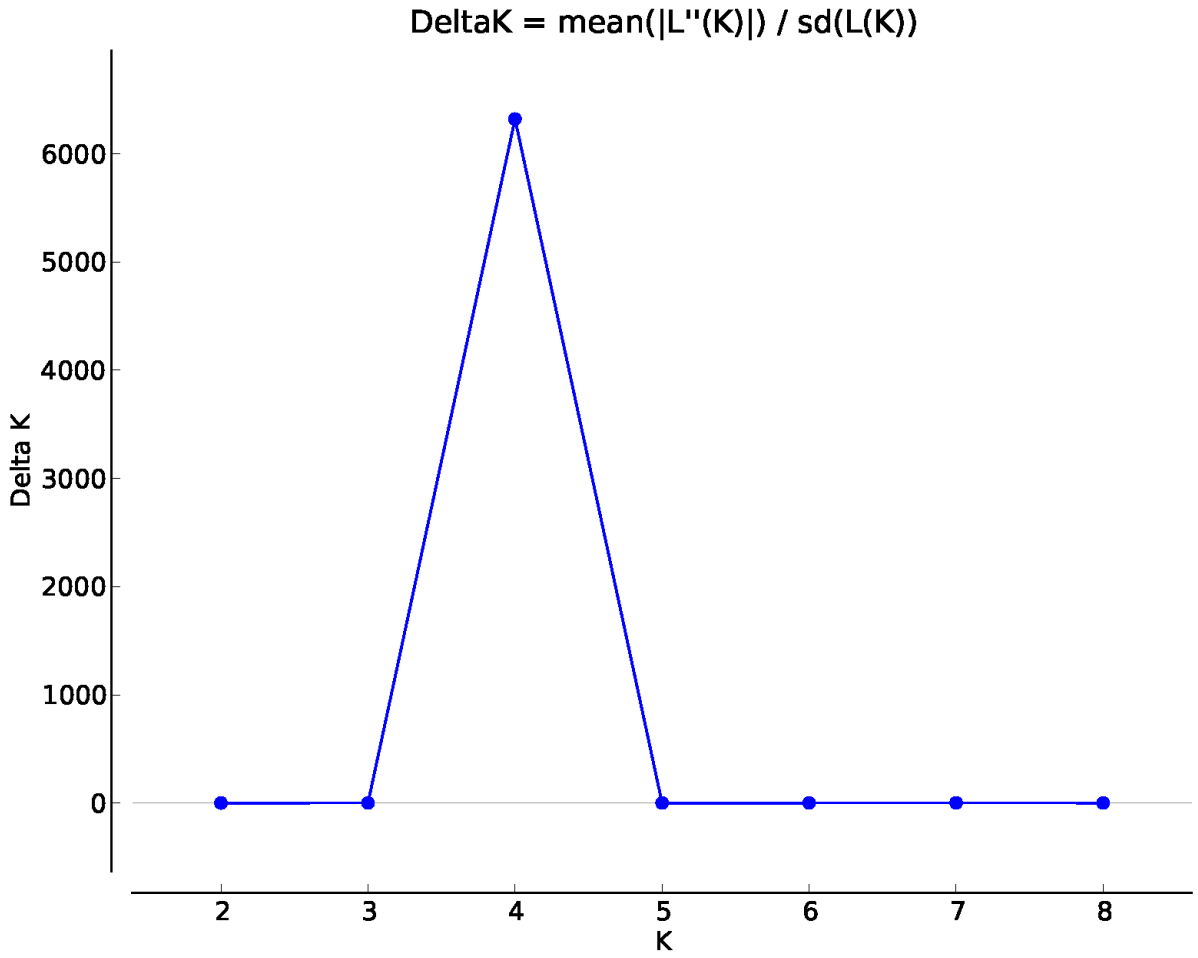
Appendix 10. Evanno ΔK values calculated for all 11 potential populations sampled (3 generations of ASL of Nile tilapia; 8 sampling locations of natural Nile tilapia, *O. niloticus*). Results are based on 3 iterations of K1-12.

APPENDIX 11

Appendix 11. Pairwise comparison of genetic distance (F_{st}) values for all three generations of the ASL and all eight natural populations of *O. niloticus* . Significant F_{st} values with a p-value < 0.05 are indicated by an asterix (*).

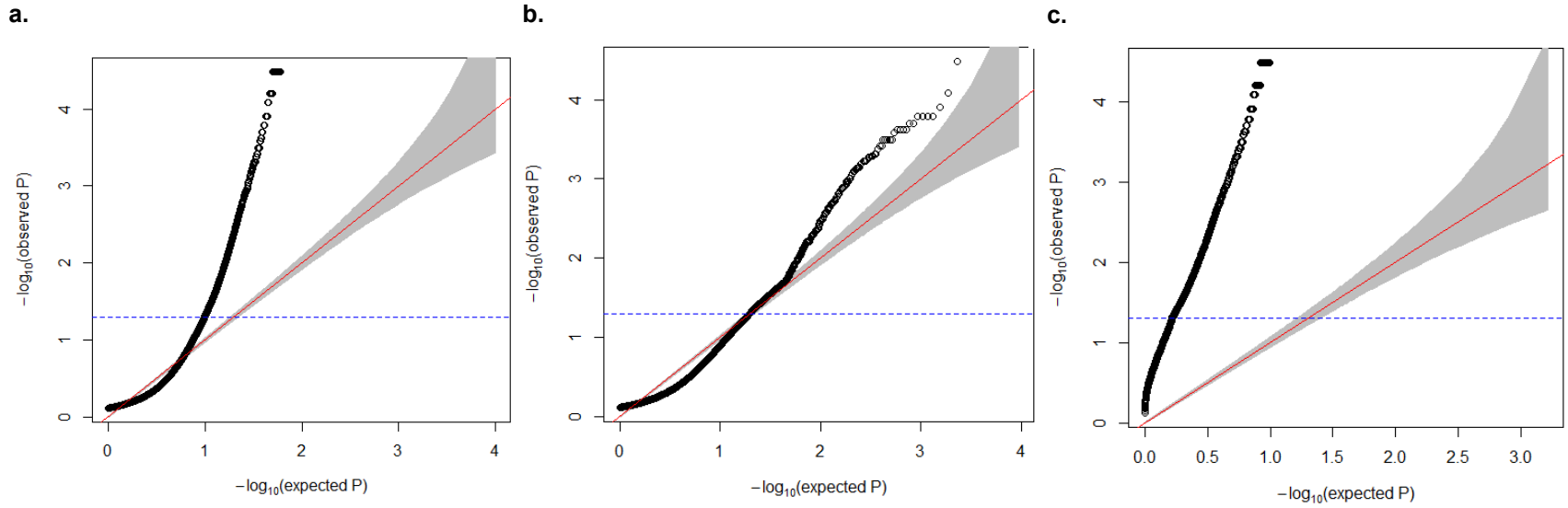
	Gen 9	***Gen 10***	***Gen 11***	***Aswan***	***Manzala Lagoon***	***Kanater***	***Lake Idku***	***Damietta***	***Lake Brulus***	***Rosetta***	***Asyut***
Gen 9	0										
Gen 10	-0.008	0									
Gen 11	-0.002	-0.004	0								
Aswan	0.045*	0.048*	0.050*	0							
Manzala Lagoon	-0.003	0.006*	0.017*	0.006*	0						
Kanater	0.035*	0.034*	0.042*	0.014*	-0.015	0					
Lake Idku	0.007*	0.011*	0.027*	0.011*	0.003*	-0.006	0				
Damietta	0.024*	0.025*	0.038*	0.020*	-0.009	0.002*	0.003*	0			
Lake Brulus	0.034*	0.035*	0.046*	0.021*	-0.013	0.006*	-0.005	0.001*	0		
Rosetta	0.024*	0.023*	0.032*	0.013*	-0.023	-0.002	-0.012	-0.009	0.000	0	
Asyut	0.055*	0.052*	0.058*	0.003*	-0.012	0.006*	-0.005	0.012*	0.014*	0.000	0

APPENDIX 12



Appendix 12. Evanno ΔK values calculated for all 8 sampling locations of natural Nile tilapia, *O. niloticus*. Results are based on 3 iterations of K1-9.

APPENDIX 13



Appendix 13. Quantile-Quantile (QQ) Plots of (a) all loci (b) neutral loci (i.e. where both balancing and directional outliers jointly identified by BayeScan v. 2.1 and Arlequin v. 3.5.2.2 were removed), and (c) $\text{neutral}_{\text{all outliers}}$ loci (i.e. where any balancing and directional outliers identified in either BayeScan v. 2.1 or Arlequin v. 3.5.2.2 were removed).The dotted blue line indicates the threshold of outliers identified at a significant ($p \leq 0.05$), the red line represent normally distributed data where the observed and expected p-value distributions are equivalent, and the surrounding grey area represents a 95% confidence interval.

APPENDIX 14

Appendix 14. All 196 outliers which could be annotated to at least one of the three (sex average, female, and male) linkage maps created in Chapter 3. Detailed below is the linkage group to which they were annotated along with the position on that linkage group on all maps to which they were annotated.

Outliers	LG	Sex Average Map Position	Female Map Position	Male Map Position
17016618|F|0--63:C>T	2	6.969		20.183
11394783|F|0--16:G>A	2	13.123		26.571
11389881|F|0--5:G>A	2	14.776	27.343	23.985
11397157|F|0--17:T>C	2	14.912		25.271
11387621|F|0--29:C>T	2	21.695	21.545	25.42
11392161|F|0--22:A>G	2	28.59	17.416	36.891
11393887|F|0--17:A>T	2	31.926	25.094	
11385511|F|0--44:T>C	2	33.367	25.314	39.536
11387012|F|0--30:C>A	2	37.303	33.85	48.766
11387020|F|0--46:T>G	2		31.299	
17021741|F|0--13:C>A	3	38.954	38.127	
11393018|F|0--23:C>G	3	40.913	38.981	
11391561|F|0--67:G>A	3	44.176		
11385225|F|0--46:A>C	3		49.538	25.099
11396389|F|0--5:G>C	4	17.013	22.692	
17019317|F|0--13:A>G	4	19.032	18.249	42.659
11391873|F|0--10:T>C	4	19.525	22.194	
11391749|F|0--56:C>T	4	20.402	24.753	
11386979|F|0--56:T>C	4	22.954	11.735	41.736
11388141|F|0--11:T>C	4	24.892	29.805	44.51
11394087|F|0--6:A>C	4	28.191	32.257	38.136
11396614|F|0--54:A>C	4	37.652	44.727	30.246

11393786|F|0--11:C>T	4	41.803	44.994	
11386226|F|0--9:G>A	4	42.92	42.31	31.347
11388966|F|0--31:C>T	4		45.989	
11385579|F|0--17:A>T	5	41.403	54.596	45.994
17888805|F|0--27:G>A	5	44.368	59.16	
11386268|F|0--23:G>A	5	62.217	74.149	26.463
11396378|F|0--26:A>C	5	64.85	63.091	23.852
11390313|F|0--9:T>G	5	65.073		24.613
11392779|F|0--59:C>G	5	66.756		20.661
17015001|F|0--8:T>G	5	68.893	79.186	
11395743|F|0--27:G>T	5		68.743	63.217
11387964|F|0--32:C>T	6	19.164	25.833	14.255
17017514|F|0--19:C>T	6	22.471	23.877	20.351
11387528|F|0--11:A>G	6	22.959	23.877	
11393141|F|0--40:C>T	6	27.668		27.108
11391838|F|0--39:C>T	6	40.323		38.791
11386442|F|0--6:G>A	6	42.337	46.166	
11389105|F|0--38:G>A	6	48.197		25.72
11387192|F|0--22:A>G	6		28.425	13.144
11396239|F|0--25:C>T	6			3.115
11396619|F|0--29:G>A	6			15.284
11397061|F|0--15:A>G	7	0		
11396443|F|0--27:C>T	7	22.174		0
11385907|F|0--68:G>A	7	28.939		13.016
11387620|F|0--54:G>T	7	45.228	25.933	34.171
11388188|F|0--59:A>G	7	58.041		
11391499|F|0--27:T>G	7	62.398		
11385680|F|0--16:G>A	7		0	18.454
11386447|F|0--17:G>A	7		0.986	

11389832|F|0--26:C>T	7		3.119	
11395983|F|0--16:T>C	7		3.185	13.046
11385224|F|0--55:T>C	7		3.388	
17012647|F|0--9:C>T	7		3.891	
11387408|F|0--60:T>C	7		3.931	
11387698|F|0--44:T>C	7		3.931	
11393767|F|0--14:C>T	7		3.931	
11394228|F|0--11:C>A	7		3.931	
11387165|F|0--50:T>A	7		4.163	
17022540|F|0--14:G>A	7		4.271	
17023948|F|0--50:C>T	7		5.675	
17006770|F|0--49:G>C	7		17.627	13.337
11386867|F|0--18:G>T	7		23.314	
11392462|F|0--32:C>T	7		28.639	41.67
17019184|F|0--35:C>T	7		49.386	62.81
11386567|F|0--24:C>T	7			5.194
17025377|F|0--24:C>T	7			18.329
11387364|F|0--44:A>G	7			19.089
11391523|F|0--51:G>A	8	6.871	20.427	
11390243|F|0--41:T>G	8	6.954		0
11387448|F|0--10:C>G	8	12.566		
11392009|F|0--22:T>C	8	19.531	16.058	16.492
11386712|F|0--16:C>T	8	38.894	49.124	39.861
11386936|F|0--12:T>G	8	40.356	45.93	34.122
11392448|F|0--25:G>A	8	43.057	42.88	39.911
11397191|F|0--25:T>C	8	44.236	43.413	41.645
11395444|F|0--61:G>C	8	46.268		36.92
11390260|F|0--16:G>A	8		39.179	9.624
11389699|F|0--34:A>G	8			39.758

11392467|F|0--47:T>A	11	48.079	70.047	
11385394|F|0--39:C>G	11	50.983	63.362	52.816
11396385|F|0--23:G>C	11	52.771	69.034	
11386014|F|0--49:G>C	11	58.572	78.823	
11387060|F|0--62:C>G	11		40.946	
11396305|F|0--38:G>A	11		55.865	41.216
11385498|F|0--36:A>G	11		65.707	
11396042|F|0--64:G>A	11		85.471	67.821
11386987|F|0--34:A>C	11			41.072
11386986|F|0--58:G>A	11			42.314
11393214|F|0--48:T>C	11			52.945
11395608|F|0--38:A>C	11			71.837
17027912|F|0--6:T>C	11			72.461
11393725|F|0--33:C>T	12	23.33	46.333	
11395235|F|0--28:C>T	12	26.582	42.749	6.053
11396518|F|0--8:C>A	12	38.041	38.529	27.824
11390063|F|0--14:G>A	12	53.068	67.362	
11388128|F|0--23:G>A	12	55.481	68.708	
11395189|F|0--54:C>T	12	56.107	67.946	39.739
11395450|F|0--48:G>A	12		68.939	47.339
11390140|F|0--15:A>G	12			8.711
11397101|F|0--60:G>A	13	28.531		46.435
11396356|F|0--5:C>G	13	33.084	24.041	
11395978|F|0--63:G>C	13	40.809		61.394
17023581|F|0--41:G>A	13	42.252	49.511	
11387951|F|0--45:A>G	13	45.686	49.547	62.783
11392586|F|0--34:T>C	13	47.162		66.658
11394693|F|0--24:A>G	13		8.57	62.007
11388983|F|0--9:G>A	13		31.595	43.233

11386691|F|0--17:T>C	13			97.139
11396558|F|0--27:C>T	14	15.848	36.528	
11388136|F|0--51:T>C	14	15.898	35.421	
11387355|F|0--32:A>G	14	22.409	24.474	
11392697|F|0--57:C>T	14	27.931		
11388231|F|0--60:A>G	14		51.28	29.374
11395896|F|0--51:C>T	14		67.337	
17019291|F|0--23:G>A	14		67.659	
17018247|F|0--54:C>T	16	12.382	0	38.375
17016550|F|0--30:A>G	16	15.928	21.268	
17023148|F|0--67:C>T	16	16.556		39.479
11391580|F|0--63:C>T	16	18.022	21.437	24.494
17013962|F|0--25:A>G	16	20.06	23.924	
17012537|F|0--56:G>T	16	20.342	24.005	32.389
11389184|F|0--27:C>G	16	20.555	22.484	
11389461|F|0--30:G>T	16	31.066	55.5	9.258
11390333|F|0--10:G>T	16	33.491	44.284	12.375
17020595|F|0--20:A>G	16	36.259	48.716	
11394102|F|0--27:C>T	16	36.847	51.96	16.385
11386596|F|0--20:G>A	16	37.531	48.591	12.219
11395435|F|0--41:G>A	16	38.151		10.234
11389268|F|0--24:G>A	16	39.5	48.638	16.008
11387902|F|0--49:C>T	16	39.509	48.421	
11387104|F|0--8:G>A	16	39.643	48.811	16.011
11396300|F|0--32:A>G	16	39.906	48.923	
11385279|F|0--44:G>A	16	40.93	50.705	16.008
11392861|F|0--23:G>A	16	41.743	52.782	16.008
11386220|F|0--36:T>A	16		33.194	
11388213|F|0--30:C>T	16		48.636	15.896

11395595|F|0--40:G>T	16		48.774	19.158
11397938|F|0--25:T>G	17	6.859	7.242	66.004
11393445|F|0--56:C>T	17	6.862	8.026	66.507
17021816|F|0--25:C>G	17	7.06	8.026	
11391597|F|0--68:T>G	17	28.432	27.65	17.963
11392333|F|0--38:C>T	17	28.535	29.07	
17888498|F|0--15:A>G	17	45.582	46.493	47.756
11394106|F|0--6:G>A	17	46.915		42.561
17012688|F|0--25:C>G	17	48.276	46.202	37.401
11396284|F|0--29:C>T	17	49.231	46.072	39.287
17881870|F|0--56:T>C	17	54.951	45.25	
11389869|F|0--25:C>G	17			52.477
11387372|F|0--43:C>T	18	48.269	39.065	25.626
17887915|F|0--7:A>C	18	54.207		29.194
11389044|F|0--18:T>G	18	65.81	52.3	7.186
17020472|F|0--10:C>T	18		54.923	22.684
17024112|F|0--9:T>C	18		56.044	30.391
11392039|F|0--52:A>G	19	9.144	3.832	52.319
11385409|F|0--59:C>G	20	50.894		32.237
11390665|F|0--10:A>G	20	50.911	54.757	16.92
11389663|F|0--14:C>T	20	51.866		
11392008|F|0--38:G>A	20	52.605	59.553	12.397
11385997|F|0--33:G>A	20	52.63	57.721	13.849
11394708|F|0--22:A>G	20	52.938		13.849
11393778|F|0--37:G>A	20	53.55	59.568	16.92
11392599|F|0--24:A>G	20	53.629	59.076	16.963
11385741|F|0--36:T>C	20	56.54	54.049	25.696
11389286|F|0--19:C>T	20	58.059	51.511	25.583
11397792|F|0--11:G>C	20	60.931	52.779	

11392400|F|0--36:G>A	20	61.538		26.23
11388195|F|0--19:T>C	20	61.667	68.38	27.899
11386198|F|0--22:C>A	20	61.692		25.379
11396310|F|0--14:T>G	20	63.213		28.563
11385372|F|0--59:C>T	20	63.91		
11394591|F|0--24:C>A	20	63.973		25.845
11392429|F|0--8:C>T	20	65.542		28.586
17024130|F|0--15:A>G	20	66.858		30.504
11389059|F|0--35:C>G	20	68.823		30.979
11387379|F|0--51:C>T	20	71.792		
11387052|F|0--33:A>G	20	73.313	73.817	
11391883|F|0--40:G>A	20	80.846		
17027530|F|0--23:C>G	20	87.291		62.443
11395473|F|0--66:G>A	20		55.3	25.583
17021064|F|0--11:A>C	20		56.965	
11393839|F|0--31:T>G	20		66.229	
11385386|F|0--25:A>G	20		69.627	
11385893|F|0--56:C>G	20			26.214
11393454|F|0--29:G>A	20			26.23
11396498|F|0--28:G>A	20			41.817
11396572|F|0--7:G>A	23	28.27	36.396	
11386860|F|0--26:G>C	23	33.74	47.874	27.281
11396584|F|0--53:C>T	23	36.379		28.522
11385953|F|0--29:A>G	23	50.683		
11387574|F|0--45:C>A	23	59.603	69.705	
11386967|F|0--22:C>T	23		35.283	29.189
11392797|F|0--68:C>A	23		62.417	27.781
11389200|F|0--38:A>G	23			26.739

APPENDIX 15

Appendix 15. Genetic diversity indicies calculated using all SNPs. Hardy-Weinberg Equilibrium (HWE) was calculated as the proporiton of markers that were significantly (p-value < 0.05) out of HWE, Monomorphic SNPs were calculated as the porportion of markers that were monomorphic, and the inbreeding coeffiient (F_{is}) was calculated per sampling location, timepoint, and/or population. Significant F_{is} values are denoted by *.

	Category	n	HWE	Monomorphic SNPs	F_{is}
Gen 9	Domestic	121	0.114	0.019	-0.035
Gen 10	Domestic	204	0.138	0.020	-0.078
Gen 11	Domestic	145	0.146	0.031	-0.054
Lake Idku	Natural	49	0.030	0.252	-0.067
Rosetta	Natural	48	0.041	0.144	0.038
Lake Brulus	Natural	50	0.032	0.263	-0.055
Damietta	Natural	50	0.041	0.213	-0.027
Manzala Lagoon	Natural	43	0.023	0.288	-0.090
Kanater	Natural	50	0.025	0.181	-0.112
Asyut	Natural	33	0.010	0.326	-0.142
Aswan	Natural	28	0.015	0.355	-0.065
Domestic		470	0.210	0.008	-0.060
Natural		351	0.194	0.057	-0.061

APPENDIX 16

Appendix 16. A subset of genetic diversity indices-including, average observed heterozygosity (H_o), expected heterozygosity (H_e), and the number of polymorphic loci- calculated using different allowances of missingness in data per sampling location and per STRUCTURE population designation (K = 2; domestic population, natural population). All loci included, all loci potentially included in analysis with loci varying per population with only 5% missingness allowed within each population (All Loci; 5% Missing Allowed Per Population), markers present in at least 50% of samples allowed (50% missingness), markers present in at least 75% of samples allowed (25% missingness), and markers present in at least 95% of samples allowed (5% missingness).

	$H_O \pm SE$					$H_E \pm SE$					Polymorphic Loci				
	All Loci	*All Loci: 5% Missing Allowed Per Population*	*50%*	*25%*	*5%*	*All Loci*	*All Loci: 5% Missing Allowed Per Population*	*50%*	*25%*	*5%*	*All Loci*	*All Loci: 5% Missing Allowed Per Population*	*50%*	*25%*	*5%*
Gen9	0.211±0.013	0.221±0.014	0.219±0.013	0.222±0.014	0.185±0.012	0.232±0.014	0.185±0.011	0.233±0.014	0.230±0.014	0.185±0.011	9,291	4519	6527	4297	1594
Gen10	0.212±0.010	0.220±0.011	0.218±0.010	0.219±0.010	0.182±0.009	0.231±0.011	0.183±0.009	0.233±0.011	0.229±0.011	0.183±0.009	8,671	3734	6497	4290	1591
Gen11	0.211±0.013	0.219±0.013	0.218±0.013	0.218±0.013	0.183±0.011	0.229±0.013	0.181±0.011	0.230±0.013	0.226±0.013	0.181±0.011	8,934	4298	6449	4257	1584
Lake Idku	0.214±0.024	0.211±0.024	0.214±0.024	0.206±0.024	0.151±0.019	0.232±0.024	0.153±0.019	0.228±0.024	0.216±0.024	0.153±0.019	6,404	3484	5143	3496	1107
Rosetta	0.181±0.029	0.178±0.024	0.181±0.024	0.271±0.028	0.115±0.020	0.212±0.029	0.133±0.018	0.209±0.024	0.253±0.028	0.133±0.018	7,626	4687	5800	2902	1368
Lake Brulus	0.221±0.024	0.214±0.024	0.221±0.024	0.216±0.026	0.153±0.020	0.236±0.024	0.154±0.019	0.232±0.024	0.211±0.024	0.154±0.019	6,754	3844	4960	3634	1045
Damietta	0.202±0.023	0.204±0.023	0.203±0.023	0.225±0.024	0.143±0.019	0.225±0.024	0.144±0.019	0.220±0.024	0.232±0.023	0.144±0.019	7,041	3752	5306	3249	1170
Manzala Lagoon	0.222±0.025	0.223±0.026	0.223±0.026	0.182±0.026	0.162±0.022	0.240±0.025	0.161±0.0210	0.237±0.025	0.205±0.025	0.161±0.021	5,942	2896	4922	3838	1037
Kanater	0.216±0.025	0.191±0.024	0.218±0.026	0.222±0.024	0.139±0.021	0.221±0.024	0.134±0.019	0.218±0.024	0.227±0.023	0.134±0.019	7,627	3591	5474	3259	1254
Asyut	0.268±0.042	0.218±0.032	0.272±0.034	0.254±0.031	0.185±0.028	0.259±0.036	0.175±0.0250	0.258±0.029	0.258±0.028	0.175±0.025	6,553	3004	4442	2858	843
Aswan	0.247±0.032	0.246±0.033	0.252±0.033	0.214±0.031	0.184±0.028	0.265±0.031	0.184±0.026	0.263±0.031	0.223±0.031	0.184±0.026	5,995	2693	4321	3380	839
Domestic	0.208±0.007	0.219±0.007	0.216±0.007	0.218±0.007	0.181±0.006	0.228±0.007	0.181±0.006	0.230±0.007	0.226±0.007	0.181±0.006	9,234	3822	6569	4326	1610
Natural	0.177±0.009	0.170±0.009	0.178±0.009	0.179±0.009	0.110±0.007	0.165±0.009	0.113±0.007	0.190±0.009	0.186±0.009	0.113±0.007	8.577	4008	6238	4098	1486