Norwegian University of Life Sciences
Faculty of Biosciences
Department of Animal and Aquacultural Sciences

Philosophiae Doctor (PhD)
Thesis 2018:101

Non-additive genetic effects in Nile tilapia

Ikke-additive genetiske effekter i Niltilapia

Rajesh Joshi

Non-additive genetic effects in Nile tilapia

Ikke-additive genetiske effekter i Niltilapia

Philosophiae Doctor (PhD) Thesis

Rajesh Joshi

Norwegian University of Life Sciences

Faculty of Biosciences

Department of Animal and Aquacultural Sciences

Ås (2018)

Thesis number 2018:101

ISSN 1894-6402

ISBN 978-82-575-1567-6

PhD Supervisors

Prof. Hans Magnus Gjøen
Department of Animal and
Aquacultural Sciences,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway.
hans.magnus.gjoen@nmbu.no

Prof. John Woolliams
The Roslin Institute, Royal (Dick)
School of Veterinary Studies,
University of Edinburgh, Easter Bush,
Midlothian, EH25 9RG, UK.
Department of Animal and
Aquacultural Sciences,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway.
john.woolliams@roslin.ed.ac.uk

Prof. Theodorus Meuwissen
Department of Animal and
Aquacultural Sciences,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway.
theo.meuwissen@nmbu.no

Dr. Matthew Kent
Department of Animal and
Aquacultural Sciences, CIGENE,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway.
matthew.peter.kent@nmbu.no

PhD Evaluation Committee

Prof. Luis Varona
Facultad de Veterinaria,
Miguel Servet, 177, 50013 Zaragoza
University of Zaragoza,
lvarona@unizar.es

Dr. Tom Druet
Unit of Animal Genomics, 1, avenue de
l'Hôpital - GIGA B34 (+1) B-4000 Liège
University of Liege
Belgium
tom.druet@uliege.be

Dr. Hanne Fjerdingby Olsen
Department of Animal and
Aquacultural Sciences,
Norwegian University of Life Sciences,
P.O. Box 5003, 1432 Ås, Norway.
hanne.fjerdingby@nmbu.no

Acknowledgements

This thesis was carried out in the Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences (NMBU) and the data used for the analysis was provided by GenoMar Genetics AS. They are acknowledged for their support.

Foremost I would like to thank my supervisors for their support, advice and guidance during my PhD. I would like to thank my main supervisor Prof. Hans Magnus Gjøen for all of his academic as well as personal help and guidance. This journey would have been much more difficult without your instantaneous help and support on many occasions. Also, I am indebted to my co-supervisor Prof. John Woolliams, who always made himself available to me despite his busy schedules, and even coming all the way from Edinburgh to Ås just to discuss my results and help me to complete my papers. I was happy that I found some similarity between John and me: that both of us liked spicy food. Your detailed comments, patience and belief in me has helped to polish my papers. Last but not least, I am equally indebted to Prof. Theodorus Meuwissen, Dr. Matthew Kent and Prof. Sigbjørn Lien for your supervision and help during my PhD.

Further, I would like to acknowledge all the co-authors, colleagues and friends at the department, who made my stay joyful. I would like to thank my office mate Gebreyohans Tesfaye, with whom I could share all the personal and professional things (and vice versa). Similarly, thanks to my other office mate Haile Wei, who used to support me most of the time when Gebreyohans and I were on some off-topic discussions. I would also thank my other professional colleagues including Solomon, Luqman, Tesfaye and Kahsay for their support and help.

During the course of my PhD I visited Hohenheim for a week. I would like to thank Gunnar for making this happen and would like to thank Jörn and Yu for their help during my stay there. During the PhD I got chance to be actively involved in SoDoC, SiN and NEPSÅ, and would like to thank all my colleagues and friends in these societies for making my stay in Norwegian winter a lot better.

My journey of PhD would have not been complete without the support my family, mainly my wife Mary Shrestha and my sister Hema Joshi. I would also thank my in-laws, family friends Saurav, Raju and their family for their help.

Ås, October 2018

Rajesh Joshi

Dedication

This thesis is dedicated to my late mother, Ms. Gaura Devi Joshi, who was always my source of inspiration and unfortunately died during my PhD after a valiant battle with cancer. This loss was a sort of compensated with the birth of my dear son, Mr. Crisper Joshi.

Table of Content

Individual papers and their supplementary have their own page numbers.

Summary

Nile tilapia (*Oreochromis niloticus*) is the world's fastest growing aquaculture species, in terms of annual increase in production (ca 10%) and is today produced world-wide. Most of the commercial and farmed Nile tilapia strains are derived from the genetically improved farmed tilapia (GIFT) strain established in the early 1990s. The systematic mixing of the eight different strains during the first 3 generations of GIFT, would prompt a hypothesis that there may have been substantial non-additive genetic effects in the base. Despite having large full-sib families, which enables the estimation of non-additive genetic effects, it is a paradox that these effects have been ignored in the commercial evaluations and the design of the breeding schemes. Thus, this thesis sets out to explore the possibility of utilizing non-additive genetic effects in Nile tilapia breeding programs using a purpose-bred population. This was achieved firstly by using classical methods utilizing pedigree to partition the variances into additive, dominance, maternal and environmental components (paper I), then by developing the necessary genomic resources (SNP-array and linkage maps) to better allow tilapia breeding to exploit new opportunities (paper II) and finally applying these resources to partition the variance components further into additive, dominance, epistasis and maternal environmental components based on marker information (paper III).

The Onil50-array for Nile tilapia containing more than 58K SNPs was developed from the whole genome sequence of 32 Nile tilapia sampled from the GenoMar breeding nucleus (paper II). The SNPs on the array were selected based on even physical distribution and the polymorphic information content. SNP performance of the array was evaluated on nearly 5000 samples, revealing high-quality genotype data for 43,588 SNPs. Then the integrated physical and genetic linkage map, containing 40,186 SNPs, was constructed. Most of the Linkage groups (LGs) were found to have sigmoid recombination pattern, with the recombination rate between males and females being 1:1.2.

A diallel design with reciprocal cross was applied to partition the variance components. The analyses by pedigree showed the presence of non-additive variation (paper I), identified to a large extent as a full-sib family component that was not associated with additive effects or maternal effects. This source is commonly assumed

to arise from dominance using pedigree. Further analysis using data from the Onil50 SNP-array (paper III) showed that this variation, when present, was associated with additive-by-additive epistasis, and not dominance. These estimates were corrected for departures from HWE. Detrimental effects of inbreeding using genomics was reported for the commercial traits of Nile tilapia.

Substantial contributions of the non-additive genetic effects were observed ($P<0.05$) for two traits: body weight at harvest (BWH) and body depth (BD). Further, substantial contribution of maternal variance ($P<0.05$) was observed for BWH, BD, fillet weight (FW) and body length (BL); estimates based on both genomics and pedigree approaches being comparable. Unlike non-additive genetic effects, including maternal component in the models was found to cause substantial consequences on the rankings for both genomic and pedigree BLUP models. Thus, ignoring maternal effect was found to inflate the heritability and introduce bias in the genetic evaluations, over-predicting the potential gains.

Rather than depending on the interactions of unknown genes contributing to the non-additive genetic variance, targeted exploitation of this effect in the future in Nile tilapia breeding program might depend on finding out the genes or genomic regions associated with the heterosis for the traits. However, the creation of maternal lines in Tilapia breeding schemes may be a possibility if this variation is found to be heritable. Though the marker information has been used in tilapia breeding for parental assignment for almost 20 years already, the genomic resources developed here have opened a new door of genomic era in Nile Tilapia breeding and has also the potential to improve the genetic gain through genomic selection.

Sammendrag

Nil-tilapia (*Oreochromis niloticus*) er den akvakulturarten som har raskest produksjonsvekst i verden, med ca 10 % årlig økning, og produseres i dag stort sett over hele verden. De fleste av de kommersielle Niltilapiastammene er kommer fra den såkalte GIFT-stammen (Genetically improved farmed tilapia) som ble etablert tidlig på 1990-tallet. Her ble åtte ulike stammer systematisk blandet i løpet av de første 3 generasjonene, noe som gir grunn til å tro at det kan ha blitt generert betydelige ikke-additive genetiske effekter i denne populasjonen. Til tross for at tilapia har store fullsøsken¬familier, som gjør det mulig å estimere ikke-additive genetiske effekter, har disse effektene hittil blitt ignorert i planlegging av avlsdesign og når en gjør avlsverdivurderinger. Derfor har denne studien valgt å undersøke muligheten for å utnytte disse ikke-additive, genetiske effektene i en Niltilapia-avlspopulasjon. Ved å bruke klassiske analysemetoder, der vi kun tar hensyn til slektskap mellom individer, beregnet vi additive-, dominans-, maternale- og miljøeffekter (artikkel 1), deretter utviklet vi en såkalt SNP-chip (SNP er en type markører) og et genetisk koblingskart, som er nyttige verktøy når en vil gjøre seleksjon i en ny art som tilapia (artikkel 2) og til slutt brukte vi disse verktøyene for å enda bedre kunne skille mellom additive-, dominans-, maternale- og miljøeffekter, som i artikkel 1, men i tillegg også additive samspillseffekter, dvs epistatiske geneffekter (artikkel 3).

Den nye SNP-chipen, Onil50, med mer enn 58 000 markører, ble utviklet spesielt for Niltilapia og er basert på full genomsekvensering av 32 Niltilapia som ble tilfeldig utvalgt fra GenoMars avlskjerne (artikkel II). Deretter ble det nye koblingskartet, sammen med et nytt fysiske genkart, konstruert basert på 40 186 SNPer. Disse ble valgt ut fra kriterier om en jevn fordeling utover kromosone og at de skulle være informative, dvs ha høy grad av polymorfi. Den nye SNP-chipen ble så testet på nesten 5000 prøver, hvilket viste at 43.588 SNPer gav genotypedata av høy kvalitet. De fleste av koblingsgruppene, som samsvarer med kromosomene, hadde et sigmoid rekombinasjonsmønster, som er forskjellig fra f.eks. det en ser hos laks. Forholdet mellom antall rekombinasjoner hos hann- og hunfisk var 1: 1,2, hvilket betyr at rekombinasjon skjer hyppigere hos hunfisk.

Et resiprokt diallelt krysningsskjema ble brukt for å kunne estimere de ønskede varianskompo-nentene. Analysene der en kun bruker slektskapsinformasjon viste en betydelig ikke-additiv genetisk variasjon (artikkel I). Denne bli i stor grad identifisert som en fullsøsken- eller familiekomponent som ikke var forbundet med additive genetiske effekter eller maternale effekter. Dette antas vanligvis å stamme fra genetiske dominanseffekter. De påfølgende analysene, ved bruk av data fra Onil50 SNP-chipen (artikkel 3) viste imidlertid at denne variasjonen, når den var tilstede, var assosiert med additive epistasieffekter, og ikke dominanseffekter. Ved hjelp av markørinformasjon kan en også estimere skadelige effekter av innavl hos Niltilapia, og vi har her for første gang påvist dette hos tilapia.

Det ble observert signifikante ($P < 0,05$) ikke-additive genetiske effektene for to egenskaper: tilvekst og kroppstykkelse. Videre ble det observert et vesentlig bidrag av maternale effekter ($P < 0,05$) for tilvekst, kroppstykkelse, filetvekt og kroppslengde. Her var estimatene basert på markørdata og slektskap sammenfallende. Det hadde videre stor betydning for rangeringen av avlskandidatene om en inkluderte maternale effekter i modellene. Dersom en ignorerende de maternale effektene vil dette gi forhøyede arvbarhetsestimater og feilaktig avlsverdier som igjen vil overestimere forventet genetisk framgangen.

I fremtiden bør en mer målrettet kunne utnytte de ikke-additive geneffektene hos Niltilapia ved å finne ut hvilke gener eller genom-områder som forårsaker heterosiseffekter for de ulike egenskapene. Opprettelsen av maternale linjer kan være en mulighet hvis denne variasjonen viser seg å være arvelig. Selv om markørinformasjonen har blitt brukt i tilapia også tidligere, vil de nye genetiske verktøyene vi her har utviklet kunne åpne en helt ny genomisk epoke for avlsarbeid med Niltilapia og vil klart gjøre det lettere å ta i bruk genomseleksjon, som for mange egenskaper vil gi en langt mer nøyaktige avlsverdiberegning.

Abbreviations

BD	Body depth
BL	Body length
BT	Body thickness
BWH	Body weight at harvest
EBV	Estimated breeding values
EEV	Estimated epistatic values
FW	Fillet weight
FY	Fillet yield
(G)BLUP	(Genomic) best linear unbiased prediction
GIFT	Genetically improved farmed tilapia
GRM	Genomic relationship matrix
GST	GenoMar supreme tilapia
HWE	Hardy Weinberg equilibrium
LD	Linkage disequilibrium
LG	Linkage group
O. niloticus	*Oreochromis niloticus*
QTL	Quantitative trait locus
RRS	Reciprocal recurrent selection
SNP	Single nucleotide polymorphism
WGS	Whole genome sequence

List of Papers

The following papers are included in the thesis and referred to in the text by their roman numbers.

I. **Joshi, R.**, Woolliams, J.A., Meuwissen, T.H.E., and Gjøen, H.M. (2018). Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity (Edinb).*, 1. doi:10.1038/s41437-017-0046-x.

II. **Joshi, R**., Arnyasi, M., Lien, S., Gjøen, H. M., Alvarez, A. T., and Kent, M. (2018). Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (O. niloticus). *Front*. *Genet*. doi: 10.3389/fgene.2018.00472.

III. **Joshi, R.**, Woolliams, J.A., Meuwissen, T.H.E., and Gjøen, H.M. (2018). Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia. *Manuscript.*

Besides these papers, I also co-authored some other papers during my PhD, which are not part of this thesis. One of them worth mentioning, which is useful for tilapia genomics is

- Conte, M. A., **Joshi, R.**, Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Roberts, R.B., Carleton, K.L., Lien. S., Kocher, T.D. (2018). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *bioRxiv*. doi:10.1101/383992.

1. General Introduction

1.1 Introduction

Nile tilapia (*Oreochromis niloticus)*, a tropical fish, is an aquaculture species farmed in over 120 countries, and is the second most important aquaculture species in the world accounting for 7.4% of global production in 2015 (Weimin, 2017) after Carp. Genetic improvement programs have contributed to the ever increasing production of Nile tilapia which is supported by more than 20 breeding programs based mainly in South East Asia (Neira, 2010). Most of the commercial and farmed Nile tilapia strains are derived from the genetically improved farmed tilapia (GIFT) base strain established in the early 1990s by crossing among four wild African strains (from Egypt, Ghana, Kenya and Senegal), and four cultured Asian strains (Israel, Singapore, Thailand and Taiwan) (Eknath *et al.*, 1993), and is considered a highly successful genetic improvement program in farmed aquaculture. The success of the scheme is demonstrated by the widespread use of GIFT in establishing aquacultural populations of tilapia, as this is suitable for both commercial and small-scale aquaculture (Worldfish, 2016). GenoMar Supreme Tilapia (GST®) strain is one of the many strains derived from GIFT, has undergone more than 25 generations of selection and is being selected for growth, fillet yield and robustness (Personal Communication, Anders Skaarud, GenoMar Genetics).

1.2 Tilapia aquaculture practice

Tilapia aquaculture ranges from subsistence farming in rural areas to large-scale commercial farming, as this species is cheap and relatively easy to produce. Nile tilapia has been shown to survive in a wide range of extensive and intensive farming systems, including both monoculture and polyculture practices in both marine and freshwater environments (Gupta and Acosta, 2004). Thus, the culture systems vary widely from cages in water-based system to ponds, raceways and tanks in land-based systems.

Nile tilapia are known to be maternal mouth brooders, where the mothers incubate the fertilized eggs in the buccal cavity. This 'brooding' is considered as care given by mother to their offspring by holding them in the mouth and is part of the normal conventional breeding practice. Normal practice in selection programs is for one

breeding male to be mated to two "ready to spawn" breeding females (nested mating design) in different breeding hapas (cages made from fine mesh) and the fertilized eggs are collected from the mouth of the females a week after spawning. These eggs are transferred to the hatching jars for incubating where they hatch 3-5 days after transfer. Then the hatched fry from each family are transferred to the separate nursery hapas, where the tagging takes place (ca. 5-7 weeks). After this, the fish are transferred to large grow-out ponds where the performance testing takes place. Fish are harvested after a grow-out period of ca. 120 days. Sexual maturity in ponds reach only at the age of 5-6 months. In tilapia, males grow faster than females and farmers prefer unisex production to avoid propagation in production ponds or cages. Hence, sex reversal using hormones (Yustiati *et al.*, 2018) is regularly practiced in commercial tilapia production. A detailed aquaculture practice of Nile tilapia is described in (Puttaraksar, 2004).

1.3 Importance of non-additive genetic effects in tilapia breeding

The history of genetics dates back to classical era (before 5th century AD) (Leroi, 2010) and has been through lots of transitions, from development of theories and models to utilizing the advanced technologies, to understand the mode of inheritance and the genetic architecture of the traits. Modern genetics can be attributed to the work of Gregor Johann Mendel in the pea-plants and his theories on genetic inheritance, commonly known as "laws of Mendelian inheritance" (Mendel, 1866). Then Ronald Aylmer Fisher took this to the next level by combining Mendelian genetics with Darwin's natural selection (Darwin, 1859) using statistical approach, which is widely known as "Theory of quantitative trait inheritance" (Fisher, 1919). To acknowledge his contribution to the quantitative genetics, various quantitative genetic societies are celebrating "the 100 years of quantitative genetics theory and its application". Fisher used the infinitesimal model to describe the resemblance between relatives in a simple additive model, which was eventually extended to incorporate the non-additive genetic effects: dominance (Wright, 1921) and epistasis (Cockerham, 1954; Kempthorne, 1954).

The basic principle of the breeding programs is to identify the best animals as the parents of the next generation. The definition of the best varies in between the

breeding programs and understanding the genetic architecture of the trait can be helpful to select the best animals based on the genetic potential and/or other criteria, define the breeding goal and maximize the genetic gain. Since it is impossible to find out the real genetic potential of the animals, we calculate the estimated genetic potential, also called the estimated breeding value (EBV) and use these EBVs to rank the animals (Falconer *et al.*, 1996).

To understand the genetic architecture of the trait, the phenotypic variance (V_P) can be partitioned into genetic (V_G), environmental (V_E) and the interaction between genetic and environmental (V_{GxE}) components (Fisher, 1919; Falconer *et al.*, 1996). Since, the tilapia aquaculture takes place in different production systems, presence of genotype × environment interaction (GxE) can be hypothesized. But it has been shown for aquaculture in broadly similar systems (e.g. ponds and cages), the GxE for the commercial traits was not significant (see the review by (Sae-Lim *et al.*, 2016) for details) and is normally assumed to be 0.

The genetic variance (V_G) can be further divided into additive (V_A) and non-additive components. The non-additive components are primarily dominance (V_D) and epistasis. Dominance variance exhibit the phenotypic deviation due to interaction of alternative alleles at a particular locus, whereas, epistatic variance exhibit the interaction between alleles at different loci (Falconer *et al.*, 1996). The first order epistatic variances are additive-by-additive (V_{AxA}), additive-by-dominance (V_{AxD}) and dominance-by-dominance (V_{DxD}) epistatic variances.

Thus, the phenotypic variance can be written as;

$$\mathbf{V_P = V_A + V_D + V_{AxA} + V_{AxD} + V_{DxD} + V_E}$$

However, there are many other possible sources for the phenotypic variances like maternal genetic effects (Mabry *et al.*, 1963), indirect genetic effects (Moore *et al.*, 1997; Muir, 2005), imprinting effects (McGrath and Solter, 1984; Surani *et al.*, 1984), cytoplasmic effects (Laipis *et al.*, 1982; Huizinga *et al.*, 1986), X-linked effects (Bulfield *et al.*, 1984) and genes affecting environmental variance (Hill and Mulder, 2010).

The variance between the sire families helps to find out the resemblance between half-sibs, which is useful in quantifying heritability. Similarly, the source of non-additive genetic effects is full-sib family variance (Falconer *et al.*, 1996). There has been various

studies to quantify the importance of the non-additive variation for various quantitative traits in agriculture and livestock (Woolliams and Wiener, 1980; Wiener *et al.*, 1992a, 1992b; Shaw and Woolliams, 1999). This has gained momentum with the development of genomics (Carlborg *et al.*, 2003, 2006, Vitezica *et al.*, 2013, 2016, 2017; Ertl *et al.*, 2014; Raidan *et al.*, 2018). In one of the early studies on non-additive influence on early growth in chickens using genomics, it was estimated that dominance and epistasis accounted for around 10% and 70% of genetic variation (Carlborg *et al.*, 2003). This example shows that the non-additive genetic effects may be an important source of genetic variations. This has been widely utilized in commercial evaluations, especially in poultry and pigs, where cross breeding is the routine practice.

The systematic mixing of the eight different strains during the first 3 generations of GIFT would prompt a hypothesis that there may have been a substantial non-additive genetic effects in the base, with reported heterosis up to 14% for body weight at harvest (BWH) (Bentsen *et al.*, 1998). Despite having large full-sib families, which enables estimation of non-additive genetic effects, it is a paradox that these effects are generally ignored in the commercial evaluations. This might be due to the complexity in the calculations involved to calculate the full non-additive genetic variations (Shaw and Woolliams, 1999), particularly with the deeper pedigree in the presence of inbreeding; and confounding between full-sib family effect with the maternal and common environmental effects. In many Nile tilapia breeding programs, this confounding is also directly related to the practice where the full-sib families are reared together until they can be pit-tagged (ca. 5-7 weeks) for identification.

Unlike the livestock species, one of the major challenges for tilapia breeding companies has been the presence of secondary markets, where the customers themselves make unauthorized utilization and distribution of the genetic materials. Thus, utilization of the non-additive genetic effects for mate selection to produce final commercial fingerlings will help to protect the genetic resources, increase the market share and increase sustainability of breeding companies without compromising the additive genetic progress.

1.4 Utilization of non-additive genetic effects in tilapia breeding programs

Ways to exploit the non-additive genetic variation was suggested in as early as 1960s [(Dickerson, 1969), see review (Sellier, 1976) and (Walsh, 2005; Carlborg *et al.*, 2006)] and this has been strongly exploited in the commercial improvement of pigs and poultry. There has been very few studies about non-additive genetic effects in aquaculture, exploring the topic within the necessary limitations imposed by aquacultural systems (Winkelman and Peterson, 1994a, 1994b; Rye and Mao, 1998; Vandeputte *et al.*, 2002; Pante *et al.*, 2002; Wang *et al.*, 2006; Gallardo *et al.*, 2010). One of the major concerns for practical utilization is the accuracy of the estimates. Even with genomics, there are difficulties in obtaining a precise estimate of the non-additive genetic effects, and most of the reported dominance and epistatic ratio in the literature have huge standard errors (e.g. (Raidan *et al.*, 2018)). The accuracy of the estimates might depend on the experimental design; and the amount and source of the available data. It has been shown that the estimation of dominance and additive-by-additive variance requires 20 and 400 times as much data required as for the additive variance (Misztal, 1997).

Like breeding values (EBVs) obtained from the standard animal model, inclusion of the non-additive genetic effects in the model also gives non-additive genetic values (values like dominance deviations or estimated epistatic values-EEVs) for the individuals. Unlike the breeding values, these non-additive genetic values are not directly transmitted to the offspring but can be utilized to obtain heterosis when appropriate crosses are made. The extra heterozygosity expressed in cross-bred offspring helps to identify the parent groups producing highest heterosis and repeated mating can be used to produce more full-sib families from that mating. The decision to select for which breeding values depends on the commercial practice: whether the commercial product is the breeding potential or the performance. For the industries marketing breeding potentials (e.g. semen in cattle) breeding value is of primary concern, whereas for the industries marketing the performance (e.g. piglets or fingerlings) heterosis expressed are of value besides the breeding values. Non-additive genetic effects have primarily been utilized in the following ways:

1.4.1. Better prediction models

There has been a continuing debate on greater prediction accuracy when the models account for the non-additive genetic effects (Lee *et al.*, 2008; Wittenburg *et al.*, 2011; Su *et al.*, 2012; Ertl *et al.*, 2014; Muñoz *et al.*, 2014; Esfandyari *et al.*, 2016; Xiang *et al.*, 2016; Piaskowski *et al.*, 2018; Raidan *et al.*, 2018). A general assertion of better prediction models can be risky, because there are lots of discrepancies of prediction accuracy, likely due to one of the following reasons:

(i) **Method used to calculate the prediction accuracy**: The results and their interpretation differs if the prediction accuracy is calculated from cross validation or from the model prediction errors. The results of the cross-validation are dependent on the training and validation sets. For example: selecting training population from all the full-sib families vs. selecting the non-related full-sib families as the training set in the same population gives different prediction accuracy. Similarly, adding more model terms in the prediction equation provides more possibility to fit the data points and might show increased prediction accuracy for the phenotype. But, with the more terms being used in estimation, the more estimation error is being introduced to a prediction.

(ii) **Definition of prediction accuracy**: Phenotypes cannot be substituted easily by breeding values when we have models with non-additive genetic effects. So, care should be taken about the definition of prediction accuracy, i.e. if they are predicting phenotypes or the breeding values.

(iii) **Use of valid tests**: Hotelling Williams tests (Dunn and Clark, 1971) are used to check the significance of the difference of two predictions when they are correlated (e.g. (Su *et al.*, 2012; Lopes *et al.*, 2015)). But care should be taken using these tests, because in cases of prediction using additive and additive-dominance models, correlations are generally very high, and these tests become invalid as the smaller differences may look like extremely highly significant.

(iv) **Genetic architecture**: If the aim is to predict the phenotype, then the genetic architecture of the trait also plays an important role in prediction accuracy (de Roos *et al.*, 2009; Morgante *et al.*, 2018). In some cases it has been seen that accounting for the genetic architecture of the traits in the models has helped to increase the

prediction accuracy of the phenotype (Morgante *et al.*, 2018), but care should be taken to generalise this.

1.4.2. Mate allocation

Mate allocations strategies are designed to influence the homozygosity or heterozygosity level of the offspring produced, which helps to manifest either inbreeding depression or heterosis. Since, the non-additive genetic effects are not generally passed on to the offspring, proper selection of mates thus can help to obtain the higher heterosis in the offspring, compared to the random mating of the parents. Mating to harvest heterosis are commonly seen in following two ways (i) mating between two distantly related lines (ii) Negative assortative mating avoiding the expression of inbreeding depression when the population is single breed.

Though the mate allocation strategies to utilize the non-additive genetic effects have been available since early days (Kinghorn, 1987; DeStefano and Hoeschele, 1992), the availability of genomics has made it much easier to predict the future performance of the offspring and the mating. The prediction of the performance of the future matings in the presence of dominance (Toro and Varona, 2010) for marker data has been given by

$$G_{ij} = \sum_{k=1}^{N} [P_{ijk}(AA)g_k + P_{ijk}(AB)d_k - P_{ijk}(BB)g_k]$$

where, G_{ij} is the predicted total genetic value of the progeny from the mating between the i^{th} and j^{th} parent; $P_{ijk}(AA)$, $P_{ijk}(AB)$ and $P_{ijk}(BB)$ are the probabilities of the genotypes *AA*, *AB* and *BB* for the combination of i^{th} and j^{th} individual and the k^{th} marker; g_k and d_k are the estimates of the additive and dominance effects of the same marker k; and N is the total number of markers.

Whereas, the predicted breeding value (a_{ij}) of the progeny is given by

$$a_{ij} = \sum_{k=1}^{N} [P_{ijk}(AA)(2 - 2p_k)\alpha_k + P_{ijk}(AB)(1 - 2p_k)\alpha_k - P_{ijk}(BB)(-2p_k)\alpha_k]$$

where, $\alpha_k = g_k + d_k(q_k - p_k)$, p_k is the minor allele frequency and $q_k = 1 - p_k$.

Hence, matings can be selected based on the total genetic value (G_{ij}) or the additive genetic effects (a_{ij}). Selection based on total genetic value maximizes the total genetic superiority, whereas the selection based on additive genetic effects maximizes the additive genetic gain. Another possibility is that the matings can be selected to maximize genetic values while maximizing additive genetic effects. This decision depends on the breeding programs and their aim; whether long-term or short-term genetic gain is desirable. In dairy breeding, the genetic value is worth nothing in current systems, whereas in pigs and chickens it may be worth a lot. However, one has to reproduce those genotypes down a multiplier chain too, which is where the line crossing is required to generate the genotypes in easy to use packages.

1.4.3. Cross breeding

Different types of breeding techniques have been used for crossbreeding to exploit heterosis (see (Mishra *et al.*, 2017) for detailed description of various categories) and are primarily popular in pig, poultry and rabbit breeding. For example, in single two breed crosses (crisscrossing as in Table 1), two breeds are crossed to produce the F_1 generation, which are then crossed to produce F_2 and so on. Heterosis decreases with the increasing number of generations. Heterosis shown by F_2 thus becomes half of that shown by F_1 (Falconer and Mackay, 1996), provided that the epistatic interactions are ignored.

Table 1: Decrease in the expected individual heterosis (H) with the increase in generations for some cross-breeding systems. Crisscrossing involves mating of two breeds x and y; rotational cross breeding involves mating of three breeds x, y and z; synthetic line mating involves mating of m breeds contributing equally to the genetic makeup of the line; and the recurrent backcrossing involves mating of x breed on x. In these crossing, some of the females are retained for replacement. Whereas in case of synthetic line, both males and females are retained for replacement. The table is adopted from (Sellier, 1976).

Generation	**Crisscrossing**	**Rotational cross breeding**	**Synthetic line mating**	**Recurrent backcrossing**
1	H_{xy}	H_{xy}	H	H_{xy}
2	½ H_{xy}	½ (H_{xz}+ H_{yz})	$(1-\frac{1}{m})H$	½ H_{xy}
3	¾ H_{xy}	¼ (H_{xy}+ $2H_{xz}$)	$(1-\frac{1}{m})H$	¼ H_{xy}
4	$^5/_8$ H_{xy}	$^1/_8$ ($2H_{yz}$+ $5H_{xy}$)	$(1-\frac{1}{m})H$	$^1/_8$ H_{xy}
5	$^{11}/_{16}$ H_{xy}	$^1/_{16}$ ($5H_{xz}$+ $9H_{yz}$)	$(1-\frac{1}{m})H$	$^1/_{16}$ H_{xy}

Different types of cross-breeding schemes have their own benefits and difficulties and the choice for the best crossbreeding approach depends on the breeding program. For example: in two breed cross breeding, the selection happens in the cross-bred progeny, but not in the breed itself. Selection in the breeds and marketing of the cross-bred offspring can be used to discontinue the loss of heterosis due to the crossing of F1. For example, reciprocal recurrent selection (RRS) helps to simultaneously improve the two populations and their crossbred offspring. Fish from two separate lines are selected randomly and mated (cross breeding) reciprocally. Sires from line 1 are mated to dams from line 2 and sires from line 2 are mated to dams from line 1 (Reciprocal cross). Then the breeding values of parents are predicted based on cross-breed performance (a type of progeny testing). The best performing males and females are selected. Since its proposition in 1949 (Comstock *et al.*, 1949), various modified RRS schemes have been practiced (Hallauer and Eberhart, 1970; Moreno-Gonzalez and Hallauer, 1982; Bouvet *et al.*, 2015), an example is shown in figure 1. Here the cross-breed selection is aided by the purebred selection. One approach to implement

this modified RRS scheme is to treat purebred performance and crossbred performance as two different traits and use selection index to get the breeding values (Wei and van der Werf, 1994).

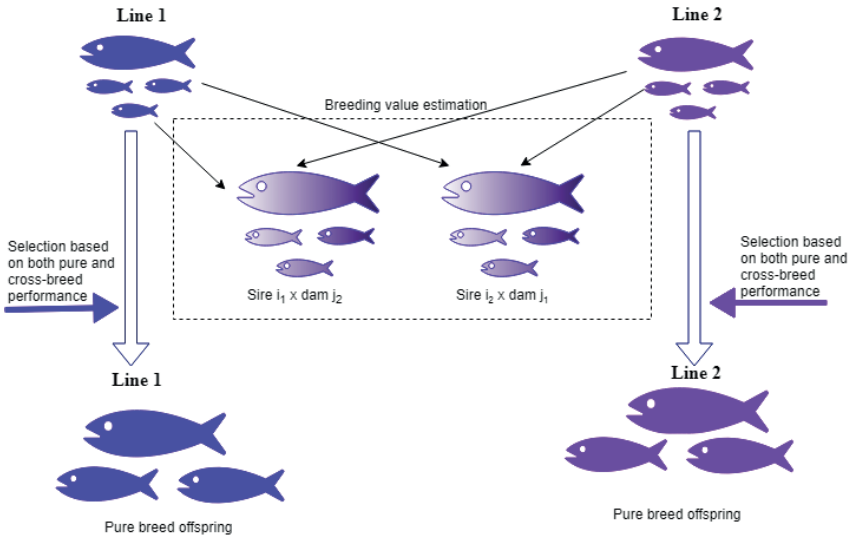


Figure 1: Modified reciprocal recurrent selection (RRS) in fish.

Genomics has offered new tools to implement cross breeding. The detailed review about non-additive genetic effects using genomics has been given by (Varona *et al.*, 2018). After calculating the SNP effects in the training population, it can be used to predict the breeding values of genotyped animal in which the phenotypes cannot be measured, e.g. for disease resistant, fillet yield, etc. Moreover, the possibility to use these predicted SNP effects over the few generations with limited loss in accuracy (Habier *et al.*, 2007; Sonesson and Meuwissen, 2009) makes it unnecessary to measure phenotypes in every generation. Hence, genomics has been used to select purebreds for crossbred performance and has shown to give greater response to selection (e.g. Dekkers, 2007; Esfandyari et al., 2015, 2016).

In general, the selection of the appropriate cross-breeding strategy not only depends on the maximizations of the annual genetic gain, but also includes the economic approach (Poutous *et al.*, 1962). So, the decision depends on the "economic-genetic optimum".

2. Aim and outline of this thesis

The aim of the thesis is to assess the possibility of utilizing non-additive genetic effects in Nile tilapia breeding programs. This aim was divided into three main sub-goals:

- To partition the variance into additive and non-additive components in Nile tilapia population designed specially to separate these components.
- To assess the impact of non-additive genetic effects on the genetic evaluation based on effects on heritability and ranking.
- To develop the genomic resources, SNP-Array and linkage map, for increased accuracy in genetic evaluation and breeding applications like genomic selection in Nile tilapia.

3. Papers

3.1 Paper I

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Heredity | doi: 10.1038/s41437-017-0046-x

Heredity

The official journal of the Genetics Society

May 2018 Vol 120 No 5
www.nature.com/hdy

SPRINGER NATURE

the geneticssociety 1919 - 2019

Heredity
https://doi.org/10.1038/s41437-017-0046-x

the geneticssociety

ARTICLE

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi[1] · JA Woolliams[1,2] · THE Meuwissen[1] · HM Gjøen[1]

Received: 29 June 2017 / Revised: 25 November 2017 / Accepted: 11 December 2017

Abstract

There are only few studies of dominance effects in non-inbred aquaculture species, since commonly used mating designs often have low power to separate dominance, maternal and common environmental effects. Here, a factorial design with reciprocal cross, common rearing of eggs and subsequent lifecycle stages and pedigree assignment using DNA microsatellites was used to separate these effects and estimate dominance (d^2) and maternal (m^2) ratios in Nile tilapia for six commercial traits. The study included observations on 2524 offspring from 155 full-sib families. Substantial contributions of dominance were observed ($P < 0.05$) for body depth (BD) and body weight at harvest (BWH) with estimates of $d^2 = 0.27$ (s.e. 0.09) and 0.23 (s.e. 0.09), respectively in the current breeding population. In addition the study found maternal variance ($P < 0.05$) for BD, BWH, body thickness and fillet weight explaining ~10% of the observed phenotypic variance. For fillet yield (FY) and body length (BL), no evidence was found for either maternal or dominance variance. For traits exhibiting maternal variance, including this effect in evaluations caused substantial re-ranking of selection candidates, but the impact of including dominance effects was notably less. Breeding schemes may benefit from utilising maternal variance in increasing accuracy of evaluations, reducing bias, and developing new lines, but the utilisation of the dominance variance may require further refinement of parameter estimates.

Introduction

Genetic variation can be partitioned into additive and non-additive components of variance, where the latter arises from the interactions among loci (epistasis) or between alleles within a locus (dominance). Although sustained genetic change in conventional breeding schemes depends only on the additive component at the time of selection, the non-additive components can be utilised in the short-term through mate selection to obtain favourable heterosis in the offspring cohort, and in the long-term to protect the genetic assets of the breeder through F2-breakdown, e.g., through selection within lines or through selection schemes like Reciprocal Recurrent Selection (RRS) (Wei and Van der Steen 1991). In practice, commercial evaluations commonly use additive models ignoring the non-additive variation, but there is a continuing debate on whether the prediction accuracy is greater when models explicitly account for the non-additive genetic variation present (Wittenburg et al. 2011; Su et al. 2012; Muñoz et al. 2014).

Relatively few studies have investigated non-additive genetic effects in fish, compared to other animals, and these are limited to few species, especially salmon (Winkelman and Peterson 1994a, b; Rye and Mao 1998; Pante et al. 2002; Gallardo et al. 2010), trout (Vandeputte et al. 2002) and carp (Wang et al. 2006), possibly due to the demands of the design for estimation. These studies have mainly been done for weight traits only, where the dominance ratio (the fraction of phenotypic variances explained by dominance deviations) ranged from 0 to 0.62. Estimates of dominance variation are lacking in tilapia, though some studies have reported heterosis effects (Bentsen et al. 1998; Maluwa and Gjerde 2006a; Lozano et al. 2011).

Electronic supplementary material The online version of this article (https://doi.org/10.1038/s41437-017-0046-x) contains supplementary material, which is available to authorized users.

✉ R Joshi
rajesh.joshi@nmbu.no

[1] Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian EH25 9RG, UK

Published online: 16 January 2018

It has been reported that the pedigree-based methods overestimate the dominance variation (Heidaritabar et al. 2016). For example, dominance and maternal effect may be confounded when analysing the data from hierarchical mating schemes (Mrode 2014); making it difficult to estimate the non-additive genetic effects precisely. In the present study, we have a factorial design with reciprocal cross, which is better suited to separate the maternal and non-additive genetic effects (Lynch and Walsh 1998; Shaw and Woolliams 1999; Vandeputte et al. 2004). The pedigree information further helps us to estimate the dominance variation by contrasting the parental dominance matrix from other effects attributed to the full-sib family groups.

The aim of this study was to study the magnitude of dominance variance, using a purpose-bred population of tilapia, on growth and morphological traits such as fillet yield. A further aim was to assess the impact on the genetic evaluation based on the effect on heritability and ranking of the selected animals.

Materials and methods

Experimental design

The data are from a trial conducted at Central Luzon State University (Munoz, Philippines) by GenoMar AS (Oslo, Norway) on Nile tilapia (*Oreochromis niloticus*) between 2014 and 2015. The test-groups studied were from the GST® strain which originated from the well-documented GIFT strain (Bentsen et al. 2017). Pedigree was thus available all the way back to the population of crossbreds defined as the base of the GIFT breeding program, which was 17 generations before the formation of the test-groups.

The mating design for the study is shown in Fig. 1a. Males and females were chosen from four full-sib families (G1, G2, G3 and G4) in generation 20, with no parents in common. From these, two parent groups were created in generation 21: group A from a G1 × G2 cross, and group B from a G3 × G4 cross. The design was intended to have 1 female parent in each of G1 and G3, and 1 male parent in each of G2 and G4, however, the offspring of G1 were subsequently found to be from 2 females, although their offspring could not be distinguished by the genotyping procedures described later. Within parent groups, 10 males and 11 females were selected from group A and 10 males and 13 females from group B. From these, A × B and B × A crosses were produced with full factorial matings across parent groups, i.e., all A females were mated with all B males, and all B females were mated with all A males. From each of these full-sib families, in Generation 22, offspring were chosen at random for rearing.

Rearing procedure

The offspring were all produced by artificial fertilisation, i.e., stripping, in three batches, which were reared separately, following the schedule of Fig. 1b. Eggs stripped from the genital papilla of ready to spawn females were fertilised in mixing containers by stripping milt from male (eggs stripped from one female was divided equally among males at 80 eggs per pool) in the wet lab and immediately transferred to incubators. There was no mouth brooding, which is common in Tilapia. To reduce the common environmental effect, the families were kept and reared together once the eggs hatched or the yolks were completely absorbed, whereas in most conventional schemes, using physical tags, families need to be kept separate until they can be tagged at a size of ca 15 g, i.e., for another 5–7 weeks. The fishes were stocked in fine-mesh nursery cages at rates of 143, 157, and 149 individuals/m^2 for batches 1–3, respectively, with corresponding survival rates of 85, 95, and 86%. All offspring were hormonally treated, so were either males or sex-reversed males, which is normal aquacultural procedure. After 21 days, tilapias were transferred to earthen grow-out ponds with stocking rate of 1.3, 1.4, and 1.5 individuals/m^2 for batches 1–3, respectively. The fish was reared under semi-intensive condition, with green-water management supplemented with commercial feed as per Genomar standard protocol (Table S1.2 and S1.3 (Supplementary 1)).

Harvesting

A total of 2987 offspring were collected after 6–7 months in the grow-out ponds, and were held or stored by batches in net cages prior to filleting, as shown in Fig. 1b. The fishes were collected smaller than normal commercial fileting size due to expected typhoon season. At collection, records were obtained for body weight (BWH), body depth (BD) and body length (BL) (Figure S1.1 (Supplementary 1)). At filleting, records of body weight (BW), body thickness (BT) and Fillet weight (FW) were obtained. Fillet yield (FY) was calculated as the ratio between fillet weight (FW) and body weight at filleting (BW) and expressed as percentage. Days of collection and filleting are shown in Fig. 1b. Batches 1 and 2 were filleted by the same three filleters, whereas batch 3 was filleted by only two of them. The data are presented in Supplementary 5.

Pedigree

Parental assignment was done by inference from 9 microsatellite markers, using DNA obtained from fin clips for parental groups A and B, and all their offspring at Temasek Life Sciences, Singapore. These microsatellites were

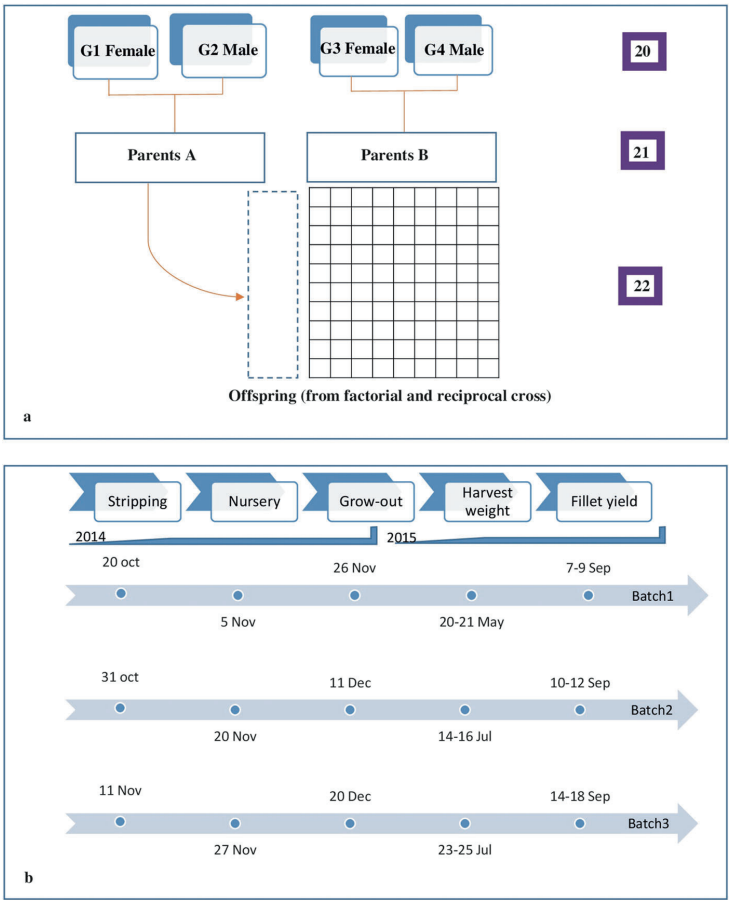


Fig. 1 **a** The mating design used for the study. The numbers on right hand side represents the generation number of the GST® strain. **b** Dates showing different phases of lifecycle of Tilapia. Offspring observed from the crosses of A and B were divided into three different batches and reared separately

selected from several hundred markers available, e.g., Kocher et al. (1998) and Lee et al. (2005), and the 9 markers chosen were all highly variable and could be run in multiplex, i.e., 9 markers in a single PCR run. The parental assignment was based on the mendelian exclusion, which is on number of hits (synonymous markers) between parent groups and offspring. It was, to minimum extent, allowed for missing genotypes or genotyping error, and the offspring having highest hit with a certain parent gets assigned to this parent (Woo-Jai Lee, personal communication).

Parentage could not be assigned for 15.6% individuals, leaving records from a total of 2524 individuals; 1318 from A × B and 1206 from B × A. These offspring were from 155 full-sib families with an average of 16.3 offspring per full-sib family (SD = 12.3, range: 1–59). The main reason for the low assignment rate is that the marker set do not have enough exclusion power for the family structure used in the cross-breeding scheme, which involved only more closely related breeders. Because of the factorial mating design, we had a lot of half-sib families, which made it harder to uniquely assign individuals. Therefore, some fish would fit equally well into 2 or more families. With no way of knowing which family was the correct one, these were set as unassigned. The complete distribution of offspring across parents and families is given in Table S1.1 (Supplementary 1).

The established pedigree from generations 3 to 22 contained 4051 records (Supplementary 6), and its structure and depth is shown graphically in Figure S1.2 (Supplementary 1). The mean inbreeding level over generations 9–18 with a mean value of 0.061 for G1–G4 is shown graphically in Figure S1.3 (Supplementary 1); being 0.061 in generation 20. The estimate of effective population size calculated

using the pedigree information from generations 9 to 18 was 95 (See Figure S1.4- Supplementary 1).

Statistical analysis

ASReml-4 (Gilmour and Thompson 2014) was used to fit mixed linear models, using REML to estimate variance components and breeding values for the six traits described above. A model with additive, dominance and maternal effects (ADM) was the full model used for analysis (see below) with dominance, maternal or both effects removed to test for their significance: sub-models AD was fitted omitting maternal effect, AM was fitted omitting dominance effect and A was fitted omitting both dominance and maternal effects. The ADM model was

$$\text{ADM Model}: \quad \mathbf{y} = \mathbf{Xb} + \mathbf{Z}_1\mathbf{a} + \mathbf{Z}_2\mathbf{d} + \mathbf{Z}_3\mathbf{m} + \mathbf{e},$$

where **y** is the vector of records; **b** is the vector of fixed effects, which were type of reciprocal cross (1 d.f.) and other systematic effects such as batch (2 d.f.) and day of collection (7 d.f.) or filleting (as appropriate, 10 d.f.); **a** is a vector of random additive genetic effects; **d** is vector of random dominance effects; **m** is vector of maternal effects; **e** is a vector of random residual errors and **X**, $\mathbf{Z}_1$, $\mathbf{Z}_2$ and $\mathbf{Z}_3$, are corresponding design matrices for fixed and random effects. For FW and FY, the fixed model also included filleter (2 d.f.)

Vectors **a** and **d** had effects for each individual in the pedigree; **m** for each full-sib family and **e** for each offspring. Their distributional assumptions were multivariate normal, with mean zero and

$$\text{Var}\begin{bmatrix}\mathbf{a}\\ \mathbf{d}\\ \mathbf{m}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{A}\sigma_A{}^2 & 0 & 0 & 0\\ 0 & \mathbf{D}\sigma_D{}^2 & 0 & 0\\ 0 & 0 & \mathbf{I}\sigma_M{}^2 & 0\\ 0 & 0 & 0 & \mathbf{I}\sigma_E{}^2\end{bmatrix},$$

where σ^2_A, σ^2_D, σ^2_M and σ^2_E are additive genetic variance, dominance genetic variance, maternal variance and error variance, respectively; **A** is the numerator relationship matrix derived from pedigree; **D** is the matrix of coefficients of fraternity for individuals in the pedigree; and **I** is an identity matrix of appropriate size. The phenotypic variance was calculated as $\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_M + \sigma^2_E$.

The estimated variance components were expressed relative to the total phenotypic variance (σ^2_P): additive heritability (h^2) $= \sigma^2_A / \sigma^2_P$, dominance ratio ($d^2$) $= \sigma^2_D / \sigma^2_P$, maternal ratio ($m^2$) $= \sigma^2_M / \sigma^2_P$. Goodness of fit was tested using likelihood ratio tests. The critical values for testing H_0: $\sigma^2 = 0$ against an alternative H_1: $\sigma^2 > 0$ with type 1 error of 0.05 was taken from the 90 percentile of $\chi_1{}^2$, i.e., 2.71.

The coefficient of fraternity between individuals x and y (Δ_{xy}) was calculated following Lynch and Walsh (1998):

$$\Delta_{xy} = \frac{\mathbf{A}_{ik} \times \mathbf{A}_{jl} + \mathbf{A}_{il} \times \mathbf{A}_{jk}}{4} \text{ for } x \neq y,$$

where i and j represents the sire and dam of x, k and l represents the sire and dam of y, $\mathbf{A}_{xy}$ is the numerator relationship between the individuals as shown in the subscripts and F is the inbreeding coefficient. For $x = y$, the coefficients were scaled by $(1-F)$ to incorporate corrections for inbreeding as per Harris (1964). The scatterplot and density plots for **A** and **D** matrix for all the individuals in the pedigree and for the phenotyped individuals are shown in Figure S1.5 (Supplementary 1). To fit the models, the inverse of **D** is required and this was calculated using the R package 'nadiv' (Wolak 2012).

Variations on this ADM model were also investigated. Firstly, the pedigree was reduced to 3 generations, treating Generation 20 as the base generation so that the estimates of h^2, m^2 and d^2 correspond more closely to a randomly mated cohort of the current population rather than the GIFT base. These were designated as A*D*M* models and procedures were identical to the ADM models other than the definition of the pedigree base.

Secondly, the analyses were conducted with a simple diallel model used to decompose the variances, which were designated SFM models (model with sire, full-sib family and maternal effects).

$$\textit{SFM} \text{ model} \qquad \mathbf{y} = \mathbf{Xb} + \mathbf{Z}_4\mathbf{s} + \mathbf{Z}_5\mathbf{m} + \mathbf{Z}_6\mathbf{f} + \mathbf{e}$$

$$\text{Var}\begin{bmatrix}\mathbf{s}\\ \mathbf{m}\\ \mathbf{f}\\ \mathbf{e}\end{bmatrix} = \begin{bmatrix}\mathbf{I}V_{\text{Sire}} & 0 & 0 & 0\\ 0 & \mathbf{I}V_{\text{Dam}} & 0 & 0\\ 0 & 0 & \mathbf{I}V_{\text{Fsib}} & 0\\ 0 & 0 & 0 & \mathbf{I}V_{\text{E}}\end{bmatrix},$$

where, the fixed effects **b** and design matrix **X** were as described for ADM models; **s** is a vector of random sire effects; **m** is a vector of random dam effects; **f** is the vector of full sib family effects; $\mathbf{Z}_4$, $\mathbf{Z}_5$ and $\mathbf{Z}_6$ are the design matrices corresponding to sire, dam and full-sib family effects. The variances attributable to the sire and dam, V_{Sire} and V_{Dam} were constrained to be equal in models S and SF models (appropriate for additive genetic contributions), with V_{Fsib} constrained to be 0 in S and unconstrained in SF. Model SM and SMF had V_{Sire} and V_{Dam} unconstrained with V_{Fsib} constrained to be 0 in SM and unconstrained in SFM. The phenotypic variance was estimated as $V_P = V_{\text{Sire}} + V_{\text{Dam}} + V_{\text{Fsib}}$. Heritabilities, maternal and dominance ratio were estimated as $h^2 = 4V_{\text{Sire}} / V_P$ and $d^2 = 4V_{\text{Fsib}} / V_P$ and $m^2 = (V_{\text{Dam}} - V_{\text{Sire}}) / V_P$.

Effects on the genetic evaluation was compared among the different models; by Pearson's correlation between

estimates of breeding values, ranking of the 100 best offspring (animals with phenotypes) and then counting the numbers that would have been excluded from the selected group compared to the simple A model.

Results

Descriptive statistics

Descriptive statistics for the six different traits are shown in Table 1. The coefficient of variation (CV) among traits ranged from 10% for body sizes (BD, BL, BT) and FY to >30% for BWH and FW.

Reciprocal cross effects

Numerical differences between reciprocal cross means were not statistically significant, although B × A were observed to have greater sizes and weights and FY; ranging from 0.1% for FY to 0.4% for BWH.

Goodness of fit

The outcomes of the likelihood ratio test for goodness of fit are presented in Table 2. The traits could be separated into three distinct groups: BL and FY showed no evidence of maternal and dominance effects; BT and FW showed evidence of maternal effects only; whereas BWH and BD showed evidence of significant maternal and dominance effects. There was direct correspondence in the significance

Table 1 Descriptive statistics for BD (cm), BL (cm), BT (mm), BWH (g), FW (g) and FY (%), where N is the number of observation, SD and SE are standard deviation and error respectively and CV is the coefficient of variation expressed as %

Traits	N	Min	Max	Median	Mean (SE)	SD	CV (%)
BD	2524	5.00	12.00	8.70	8.86 (0.02)	1.00	11
BL	2524	14.10	28.20	22.40	22.37 (0.04)	2.14	10
BT	2513	23.50	59.70	40.40	40.65 (0.09)	4.40	11
BWH	2524	107.80	804.70	385.70	403.83 (2.48)	124.82	31
FW	2524	16.20	342.60	134.50	141.51 (1.02)	51.37	36
FY	2513	12.12	54.67	33.01	32.64 (0.06)	3.19	10

Table 2 Log likelihood values of different models for the six traits

Model	BD	BL	BT	BWH	FW	FY
ADM models						
A	−81.786	−90.60	−56.94	−28.34	−21.95	−68.37
AD	-79.494^{+}	−90.34	−56.14	-26.66^{+}	−21.46	−68.37
AM	-78.220^{+}	−89.38	-55.55^{+}	-24.15^{+}	-15.59^{+}	−68.33
ADM	$-75.660^{\uparrow}$	−89.05	−54.57	$-22.19^{\uparrow}$	−14.80	−68.33
*A*D*M* models*						
A*	−81.79	−90.61	−56.94	−28.33	−21.94	−68.36
A*D*	-79.50^{+}	−90.36	−56.14	-26.66^{+}	−21.45	−68.36
A*M*	-78.23^{+}	−89.38	-55.55^{+}	-24.15^{+}	-15.60^{+}	−68.32
A*D*M*	$-75.65^{\uparrow}$	−89.05	−54.58	$-22.18^{\uparrow}$	−14.83	−68.32
SFM models						
S	−81.787	−90.61	−56.94	−28.34	−21.95	−68.37
SF	-79.498^{+}	−90.36	−56.14	-26.66^{+}	−21.45	−68.37
SM	-78.225^{+}	−89.38	-55.55^{+}	-24.15^{+}	-15.59^{+}	−68.33
SFM	-75.645^{-}	−89.05	$-54.57^{\uparrow}$	-22.17^{-}	-14.83^{-}	−68.33

In animal models, superscripts +, − and ↑ are used to denote significance tests (LRT) within the hierarchy of models. Superscript + indicates significance over model A, and ↑ indicates significance over A, AD and AM models. Similarly, in Sire and Dam models, + indicates significance over model S, ↑ indicates significance over S and SF models, and − indicates significance over S, SF and SM models

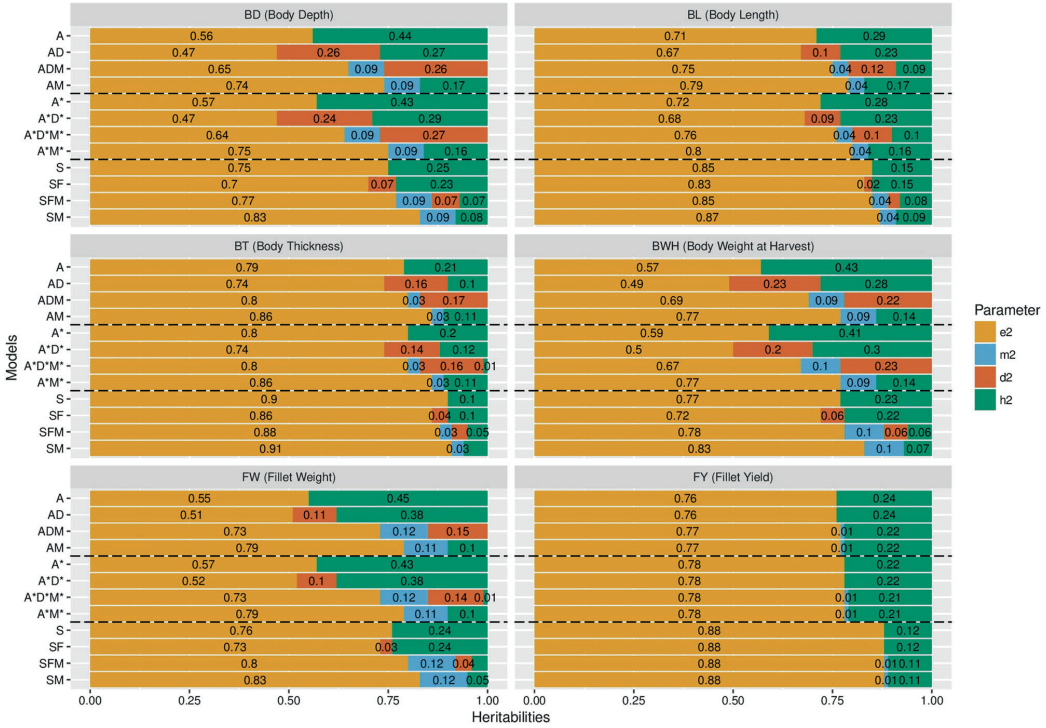


Fig. 2 Decomposition of phenotypic variance into additive (h^2), dominance (d^2), maternal (m^2) and residual (e^2) components for the six traits studied. Missing values of m^2 for some model means that the values are similar to the values obtained from other models for same trait. A was fixed to zero in the ADM model for all traits except BL, D was fixed to zero in both the AD and the ADM models for the trait FY, and F was fixed to zero or was in borderline in the SFM and SF models for FY

Table 3 Heritabilities and phenotypic variances for the models of best fit for different traits (SE are in parentheses)

Traits	SFM models				A*D*M* models			ADM models		
	h^2	d^2	m^2	σ^2_P	h^2	d^2	m^2	h^2	d^2	m^2
BD	0.07 (0.04)	0.07 (0.04)	0.09 (0.04)	0.51 (0.03)	0 (0)	0.27 (0.09)	0.09 (0.04)	0 (0)	0.26 (0.09)	0.09 (0.04)
BWH	0.06 (0.03)	0.06 (0.04)	0.10 (0.04)	6681 (355)	0 (0)	0.23 (0.09)	0.10 (0.04)	0 (0)	0.22 (0.08)	0.09 (0.04)
BT	0.06 (0.03)	—	0.03 (0.02)	8.89 (0.31)	0.11 (0.06)	—	0.03 (0.02)	0.11 (0.06)	—	0.03 (0.02)
FW	0.05 (0.03)	—	0.12 (0.05)	1002 (58)	0.10 (0.05)	—	0.11 (0.05)	0.10 (0.05)	—	0.11 (0.05)
BL	0.15 (0.05)	—	—	3.00 (0.11)	0.28 (0.08)	—	—	0.29 (0.08)	—	—
FY	0.12 (0.04)	—	—	8.95 (0.29)	0.23 (0.07)	—	—	0.24 (0.07)	—	—

of these sources of variation (dominance and maternal) across the classes of model ADM, A*D*M* and SFM. This is explained in Supplementary 2.

Estimates of heritabilities

Estimates for the variance components and heritabilities for different traits obtained by the different models are shown graphically in Fig. 2 and in detail in Table S3.1 (Supplementary 3). The summary of the models of best fit for all the traits are given in Table 3.

The simple models gave the greatest additive genetic variances, and greatest h^2 for all traits. The inclusion of dominance in the models decreased the additive variance in ADM and A*D*M* models but only marginally in SFM models. In contrast, including maternal effect decreased the

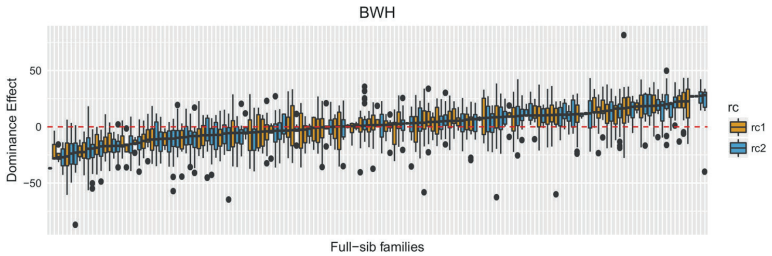


Fig. 3 Figure showing the boxplots of the dominance deviations for individuals in different full-sib families obtained from ADM models for BWH (g). Boxplots are colour coded for the reciprocal crosses

additive genetic variance considerably for some traits. ADM and A*D*M* models gave similar results for all the traits.

For BWH and BD, the two traits for which the best fit included dominance, the dominance ratio was found to be 0.06 ± 0.04 and 0.07 ± 0.04 using the SFM model, but was much greater, with corresponding greater standard errors, for ADM and A*D*M* models; 0.27 ± 0.09 and 0.23 ± 0.08, respectively for the A*D*M* models. The dominance deviation among and within the different full-sib families for BWH are presented in Fig. 3, indicating large differences in expressed dominance effects.

For the four traits where evidence of maternal ratio (m^2) was found ($P < 0.05$), the fraction was close to 0.1 for FW, BD and BWH; but was smaller for BT.

As shown in Fig. 2, h^2 for all traits other than FY depend heavily on the model fitted. For best fit A*D*M* models, the estimates of h^2 were moderate for BL and FY (0.28 and 0.23, respectively) which showed only additive variation; small for BT and FW (0.11 and 0.1, respectively), where there was evidence of maternal effects but no dominance, and 0 for BD and BWH, which showed both dominance and maternal variation. In the latter case, estimates of h^2 from SFM models were small (0.07 ± 0.04 and $0.06 \pm / 0.03$ for BD and BWH, respectively) rather than 0.

Table 4 The impact of model choice for the top 100 animals after ranking animals on EBVs for the six traits compared to a model fitting only additive genetic variance or an A model

Models	Comparison based on A model					
	BD	BWH	BT	FW	BL	FY
A	0	0	0	0	0	0
AD	6	1	4	1	1	0
ADM	—	—	—	—	48	5
AM	52	52	26	53	43	5
A*	0	0	0	0	0	0
A*D*	7	2	4	1	1	0
A*D*M*	—	—	30	58	48	5
A*M*	52	52	26	53	43	5

The number shown is the number of top animals in A models that are excluded when fitting an alternative model, therefore the 0 for the A model is by definition. The dash indicates no additive variation was detected and so no EBVs were available

Change in ranking

The difference in ranking of Estimated Breeding Values (EBVs) among the 100 best animals, as a result of different models and the use of different depth of pedigree, is presented in Table 4, for which the cohort using the simple A model has been used as a reference group for each trait. Adding only dominance effect made only minor differences in the top 100 list, with only 1–6% of the animals changing across the various traits. In contrast including maternal effect changed ~50% of the animals in the list for traits where best fit models indicated maternal variance, with much smaller impacts for BT and FY, where the maternal effect was not statistically significant. There was very little difference between ADM and A*D*M* models, showing the change of base from generation 3 to 20 had little impact. SFM models are not shown in Table 4 as these do not provide estimates of EBV.

Correlation of the EBVs

The correlation of the EBVs for all animals with observations among the different models are presented in Fig. 4. The correlations were close to 1 when only dominance term was added, i.e., changing from A to AD or AM to ADM or the analogous changes in A*D*M' models, which is consistent with the outcomes of the ranking shown in Table 4.

However, including the maternal effects was found to be different for different traits, ranging from 0.76 for FW to 0.94 for BL. For FY, the correlation was 1, as the maternal effect was close to 0. As with the ranking, there was little impact from changing the base of generation 3 (ADM models) to generation 20 (A*D*M* models)

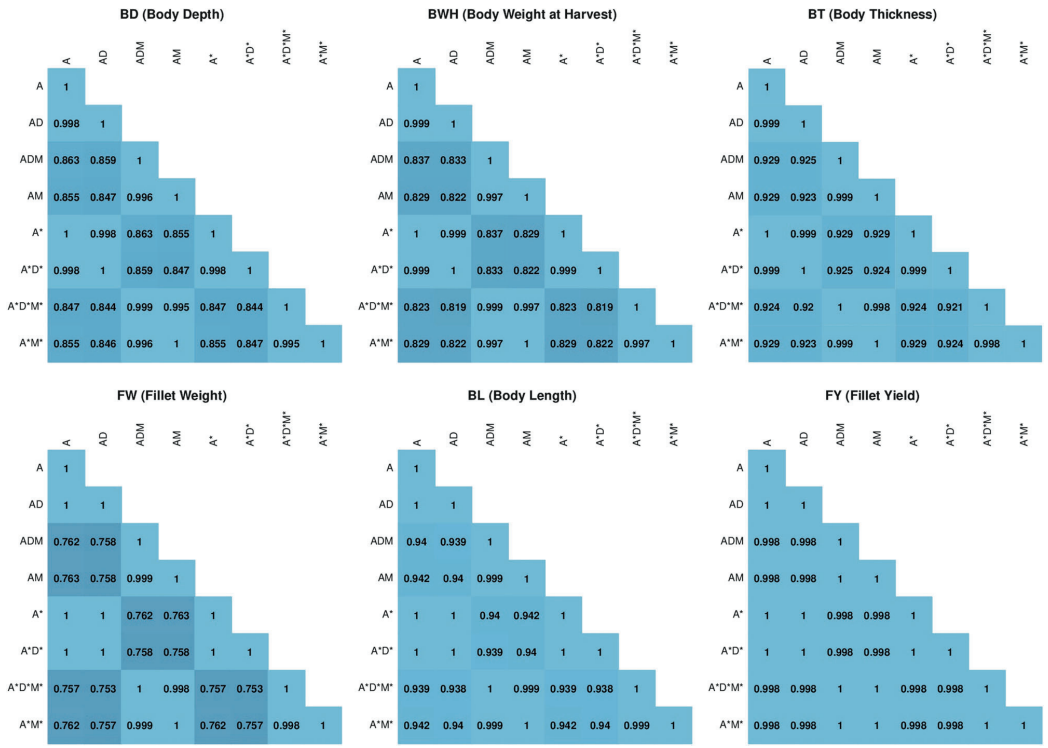


Fig. 4 Correlation values of the EBVs for different models and traits (The colour has been coded from dark to light blue, signifying low correlation for darker colours). Please note: A was fixed to zero in the ADM model for all traits except BL, D was fixed to zero in both the AD and the ADM models for the trait FY, and F was fixed to zero or was in borderline in the SFM and SF models for FY

Discussion

To the best of our knowledge, these are the first published results on dominance ratios in tilapia, and are potentially important for commercial production, both for the accuracy of EBVs for use in selection and for the eventual utilisation of heterotic effects. These were obtained by separating out the additive and non-additive genetic effects from the maternal and common environmental effects. This was achieved using a factorial mating design, including reciprocal crosses, and exploiting the large full-sib family sizes possible in fish species, which is uncommon in livestock, and seldom used in commercial aquaculture. The scope of the trial encompassed both the commercially important morphological and weight-related traits, and the post-harvest measures of fillet weight and yield, which characterise the primary saleable product.

The GST® strain used for this study is derived from the GIFT strain that is the common ancestor to most tilapia populations used for commercial breeding, and the first 10 generations of GST® also correspond to the first 10 generations of the GIFT strain. The designated base generation of GIFT, which here is defined as generation 3 of GST®, was formed from four wild and four Asian strains crossed systematically over 3 generations to allow mixing of the strains before selection for growth was commenced (Eknath et al. 1993). This origin from several diverse strains would prompt a hypothesis that there may have been substantial non-additive genetic variation in this base. The heterosis between different pairs of founding strains was reported range from <1 to 14% for BWH (Bentsen et al. 1998). For *Oreochromis shiranus*, a different tilapia species, the heterosis between strains in F1 crosses was up to 15% for BWH (Maluwa and Gjerde 2006b). The continued existence and the magnitude of the initial non-additive variation in the current GST® strain would be subject to the changes in the frequencies of alleles underlying this variation, and the partition between dominance and additive variation will change over time accordingly (Falconer et al. 1996). Estimation of the base variances using the ADM linear mixed models does not account for these changes in allele frequency.

Source of dominance variation

This variance parameters obtained from the ADM, A*D*M* and SFM models are all interpretations of the same three core variance components that are intrinsic to the factorial design, as shown in Supplementary 2. These core components are the variances among sires (V_{Sire}), variances among dams (V_{Dam}), and the variances within full-sib families (V_{Fsib}). The supplementary information (Supplementary 4) shows that projecting V_{Fsib} to estimate σ^2_D in the GST® base generation results in a 4.5-fold scaling of the value that would be obtained from a standard assumption that V_{Fsib} is $¼\sigma_D{}^2$. This explains why a small variance component in SFM models can translate into substantial estimates of σ^2_D in ADM models. Furthermore, estimates of σ^2_A from V_{Sire} are influenced by the design in that the sires used within parent groups A and B are full sibs. Therefore, the models produce a range of estimates that might be considered: empirical SFM estimates assuming $\sigma^2_A = 4V_{Sire}$, $\sigma^2_M = (V_{Dam} - V_{Sire})$, if >0; and $\sigma^2_D = 4V_{Fsib}$; A*D*M* estimates with a base generation in generation 20, which most closely correspond to random mating in the current population; and ADM estimates which project back to generation 3, the GIFT base. Since each emerge as scaling of the same set of core components, the standard errors and uncertainties reflect the magnitude of the scaling factors applied. The near-equal scaling factors from using generation 3 (ADM) or 20 (A*D*M*) as the base, demonstrate that the scaling observed for estimates of σ^2_D in ADM models is a consequence of the design rather than the additional pedigree. There are additional approximations in the use of the fraternity matrix to assess dominance, as it is an approximation of the full dominance model (for example, Shaw and Woolliams (1999)), and it excludes terms that increase in importance with the inbreeding coefficient, F. The relatively low value of F suggests this may not be a serious problem in ADM models, and for A*D*M models with a generation 20 base, where $F = 0$.

Estimates of different variance components

It has been assumed that V_{Fsib} can be interpreted as dominance variance, an assumption common to many other studies. Although our design has separated out the maternal effect and minimised common environmental effects through the management described in the Materials and Methods, this interpretation cannot be certain. The results show that maternal variance is still detectable for four of the six traits (not for BL and FY) despite this management. These effects might be related to the size and quality of the eggs or mitochondrial effects. Large eggs have more yolk reserves and have been shown to be positively correlated to the growth and development of fry (Rana 1985; Springate and Bromage 1985). There has been no separate reporting of maternal ratio in the tilapia studies listed in Table S3.2 of Supplementary 3, since their design did not allow to separate them from common environment or full-sib family effects (e.g. GIFT has a hierarchical mating design).

The estimated h^2 for all traits, except for BWH and FW, are within the ranges of those published for GIFT (Table S3.1 of Supplementary 3) although for BWH and FW our estimates are towards the low end of the range. One contributing reason for this is that we have used the complex models which will have removed maternal and full-sib variances that may have been miss-attributed in simpler models. For example, the best-fit estimate of h^2 for BWH, which tends to be particularly low in comparison to other estimates from GIFT or GST®, is 0.40 for the A* model, which is similar to the other published estimates for GIFT and GST®. However, the low heritability estimates reported in the present study, must be evaluated as too low, since the realised genetic gain found in many tilapia studies, e.g., as reported by Bentsen et al. (2017). On the other hand, such high selection response is expected in the initial phase of a breeding program, since considerable "Bulmer effect" will cause higher selection response than in later phases of the selection program (Bulmer 1971). The correct heritability estimates thus probably will be somewhere between these boundaries.

There have been no previous estimates of dominance ratios in tilapia, but very few in other fish species, including the more intensively studied trout and salmon (Winkelman and Peterson 1994a, b; Rye and Mao 1998; Pante et al. 2002; Gallardo et al. 2010), with moderate values of the dominance ratio for BWH. But the comparison is not straightforward, the mating designs in these studies have low power to separate the dominance and common environment effect; the source for the dominance variation being only from the multi-generational pedigree, with phenotypes available at each generation. The significance and the standard errors of the d^2 were not reported for Atlantic salmon (Rye and Mao 1998); and d^2 was not significantly different from zero for chinook salmon (Winkelman and Peterson 1994a, b). Although d^2 was stated as significant (0.19 and 0.06 for two populations, but with no s.e.) for coho salmon (Gallardo et al. 2010), they were unable to separate dominance from common environment precisely with their applied mating and rearing design.

Implications for aquaculture production

The maternal component, shown to be present in all but two of the traits, has practical consequences for the genetic evaluation. This source of variance is not always fitted, however, as shown in Table 4, it can have substantial consequences on the rankings of selection candidates.

Furthermore, ignoring the term will tend to inflate the heritability and consequently introduce bias into evaluations; over-predicting potential gains. It also places importance on management steps to minimise the size of this component, although, as yet it may not be feasible to remove the component completely from all traits, as demonstrated in this study. The finding that some traits exhibit dominance variance will likely require further research as its magnitude remains uncertain and obtaining further information remains challenging, although genomics may offer new opportunities because high-density SNP genotypes provide more individual genomic information, potentially leading to more accurate estimate of the relationships and dominance variance (Vitezica et al. 2013; Heidaritabar et al. 2016). Including dominance, terms when parameters are open to substantial error may reduce the accuracy of prediction rather than improve it (Sales and Hill 1976). Furthermore, adding dominance to models had little impact on ranking the EBVs in this study and have had only marginal benefit in other sectors (e.g. Sun et al. 2014). However, the findings do open consideration of specialised breeding options. The maternal variance may be heritable, and instead of minimising it there may be opportunities for breeding specialised maternal and sire lines to breed crossbred fish using reciprocal recurrent selection, which could become more attractive if further research confirms the existence of substantial dominance variance in commercially important traits. This may also involve the utilisation of the relatively large differences in expressed dominance effects among and within families, as shown in Fig. 3.

Data Archiving

The phenotypic data and the pedigree are available as Supplementary 5 and 6 respectively.

Acknowledgements We would like to thank Genomar AS for providing the data, in particular Anders Skaarud, Alejandro Tola Alvarez and Marietta Palada De Vera. ASReml analysis was performed in the Abel Cluster, Oslo.

Author contributions HMG conceived and designed the study, RJ did the statistical analysis, JAW contributed to this analysis, and all authors contributed to the discussion of the results and writing of the paper.

Compliance with ethical standards

Conflict of interest The authors declare that they have no competing interests.

References

Bentsen HB, Eknath AE, Palada-de Vera MS, Danting JC, Bolivar HL, Reyes RA et al. (1998) Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of Oreochromis niloticus. Aquaculture 160:145–173

Bentsen HB, Gjerde B, Eknath AE, de Vera MSP, Velasco RR, Danting JC et al. (2017) Genetic improvement of farmed tilapias: Response to five generations of selection for increased body weight at harvest in Oreochromis niloticus and the further impact of the project. Aquaculture 468:206–217

Bulmer MG (1971) The effect of selection on genetic variability. Am Nat 105:201–211

Eknath AE, Tayamen MM, Palada-de Vera MS, Danting JC, Reyes RA, Dionisio EE et al. (1993) Genetic improvement of farmed tilapias: the growth performance of eight strains of Oreochromis niloticus tested in different farm environments. Aquaculture 111:171–188

Falconer DS, Mackay TF, Frankham R (1996) Introduction to quantitative genetics (4th edn). Trends Genet 12:280

Gallardo JA, Lhorente JP, Neira R (2010) The consequences of including non-additive effects on the genetic evaluation of harvest body weight in Coho salmon (Oncorhynchus kisutch). Genet Sel Evol 42:19

Gilmour A, Thompson R (2014) ASReml 4 Australasian Statistics Conference 2014. Statistical Society of Austraila:Post Lincoln, 2014.

Harris DL (1964) Genotypic covariances between inbred relatives. Genetics 50:1319

Heidaritabar M, Wolc A, Arango J, Zeng J, Settar P, Fulton JE, et al. (2016) Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. J Anim Breed Genet 133:334–346.

Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B (1998) A genetic linkage map of a cichlid fish, the tilapia (Oreochromis niloticus). Genetics 148:1225–1232

Lee B-Y, Lee W-J, Streelman JT, Carleton KL, Howe AE, Hulata G et al. (2005) A second-generation genetic linkage map of tilapia (Oreochromis spp.). Genetics 170:237–244

Lozano C, Gjerde B, Bentsen HB, Dionisio EE, Rye M (2011) Estimates of strain additive genetic, heterosis and reciprocal effects for male proportion in Nile tilapia, Oreochromis niloticus L. Aquaculture 312:32–42

Lynch M, Walsh B (1998) Genetics and analysis of quantitative traits. Sinauer Associates, Inc. Publishers, Sunderland, MA.

Maluwa AO, Gjerde B (2006a) Estimates of the strain additive, maternal and heterosis genetic effects for harvest body weight of an F 2 generation of Oreochromis shiranus. Aquaculture 259:38–46

Maluwa AO, Gjerde B (2006b) Genetic evaluation of four strains of Oreochromis shiranus for harvest body weight in a diallel cross. Aquaculture 259:28–37

Mrode RA (2014) Linear models for the prediction of animal breeding values. CABI, Wallingford, Oxfordshire, UK

Muñoz PR, Resende MFR, Gezan SA, Resende MDV, de Los Campos G, Kirst M et al. (2014) Unraveling additive from nonadditive

effects using genomic relationship matrices. Genetics 198:1759–1768

Pante MJR, Gjerde B, McMillan I, Misztal I (2002) Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, Oncorhynchus mykiss. Aquaculture 204:383–392

Rana KJ (1985) Influence of egg size on the growth, onset of feeding, point-of-no-return, and survival of unfed Oreochromis mossambicus fry. Aquaculture 46:119–131

Rye M, Mao IL (1998) Nonadditive genetic effects and inbreeding depression for body weight in Atlantic salmon (Salmo salar L.). Livest Prod Sci 57:15–22

Sales J, Hill WG (1976) Effect of sampling errors on efficiency of selection indices 1. Use Inf Relat Single Trait Improv Anim Prod 22:1–17

Shaw FH, Woolliams JA (1999) Variance component analysis of skin and weight data for sheep subjected to rapid inbreeding. Genet Sel Evol 31:1

Springate JRC, Bromage NR (1985) Effects of egg size on early growth and survival in rainbow trout (Salmo gairdneri Richardson). Aquaculture 47:163–172

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS (2012) Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS ONE 7:e45293

Sun C, VanRaden PM, Cole JB, O'Connell JR, Reinhardt F (2014) Improvement of prediction ability for genomic selection of dairy cattle by including dominance effects (W Barendse, Ed.). PLoS ONE 9:e103934

Vandeputte M, Kocour M, Mauger S, Dupont-Nivet M, De Guerry D, Rodina M et al. (2004) Heritability estimates for growth-related traits using microsatellite parentage assignment in juvenile common carp (Cyprinus carpio L.). Aquaculture 235:223–236

Vandeputte M, Quillet E, Chevassus B (2002) Early development and survival in brown trout (Salmo trutta fario L.): indirect effects of selection for growth rate and estimation of genetic parameters. Aquaculture 204:435–445

Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195:1223–1230

Wang C, Li S, Xiang S, Wang J, Liu Z, Pang Z et al. (2006) Additive, dominance genetic effects for growth-related traits in common carp, Cyprinus carpio L. Aquac Res 37:1481–1486

Wei M, Van der Steen HAM (1991) Comparison of reciprocal recurrent selection with pure-line selection systems in animal breeding (a review). Anim Breed Abstr 59:281–298. In:

Winkelman AM, Peterson RG (1994a) Genetic parameters (heritabilities, dominance ratios and genetic correlations) for body weight and length of chinook salmon after 9 and 22 months of saltwater rearing. Aquaculture 125:31–36

Winkelman AM, Peterson RG (1994b) Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in chinook salmon. Aquaculture 125:17–30

Wittenburg D, Melzer N, Reinsch N (2011) Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers. BMC Genet 12:74

Wolak ME (2012) nadiv: an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. Methods Ecol Evol 3:792–796

Supplement Information 1

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi[1], J Woolliams[1,2], THE Meuwissen[1] and HM Gjøen[1]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] The Roslin Institute, Royal (Dick) School of Veterinary Studies, The University of Edinburgh, Easter Bush Campus, Midlothian, Scotland, United Kingdom

Design of the study

Table S1.1: Observations in each factorial mating. 18 different sires and dams are mated in factorial manner

A x B	**S1**	**S2**	**S3**	**S4**	**S5**	**S6**	**S7**	**S8**	**S9**	**Total**
D1	5	-	3	5	1	4	2	-	-	**20**
D2	9	-	6	3	2	2	6	4	-	**32**
D3	30	6	46	24	13	36	14	17	14	**200**
D4	5	8	8	9	2	4	-	4	6	**46**
D5	5	7	4	3	1	3	-	6	3	**32**
D6	26	20	39	13	8	10	16	17	12	**161**
D7	26	12	44	13	13	30	13	17	22	**190**
D8	34	25	59	19	22	47	15	34	15	**270**
D9	15	10	17	8	8	10	12	7	8	**95**
D10	35	27	54	16	21	52	45	4	18	**272**
Total	**190**	**115**	**280**	**113**	**91**	**198**	**123**	**110**	**98**	**1318**

B x A	**S10**	**S11**	**S12**	**S13**	**S14**	**S15**	**S16**	**S17**	**S18**	**Total**
D11	30	29	13	32	29	35	19	16	22	**225**
D12	20	18	7	9	16	27	3	15	13	**128**
D13	18	25	19	16	26	16	14	22	17	**173**
D14	11	6	6	8	9	8	3	6	4	**61**
D15	33	36	23	32	37	38	19	28	26	**272**
D16	9	8	11	7	11	17	7	12	15	**97**
D17	3	13	6	2	3	6	1	4	3	**41**
D18	16	36	14	30	27	30	10	29	17	**209**
Total	**140**	**171**	**99**	**136**	**158**	**177**	**76**	**132**	**117**	**1206**

Table S1.2: Minimum and maximum temperature during tilapia production phase

Month	Temperature (°C)	
	Min	Max
Nov. 2014	28	31
Dec. 2014 (1-15)*	27	30
Dec. 2014 (16-31)*	25	29
Jan. 2015	25	28
Feb. 2015	25	28
Mar. 2015	27	30
Apr. 2015	27	31
Aug. 2015	28	31
Sep. 2015	30	32

*Numbers inside the parenthesis indicate days.

Table S1.3: Types of commercial feed fed during different tilapia production stages

Size of fish	Type
AI/fry	Booster1
Nursery	Booster1 & 2
1-10 g	Booster3
11-20 g	Pre-starter
20-80 g	Starter
81 g & above	Grower
Breeders	Broodstock/Grower

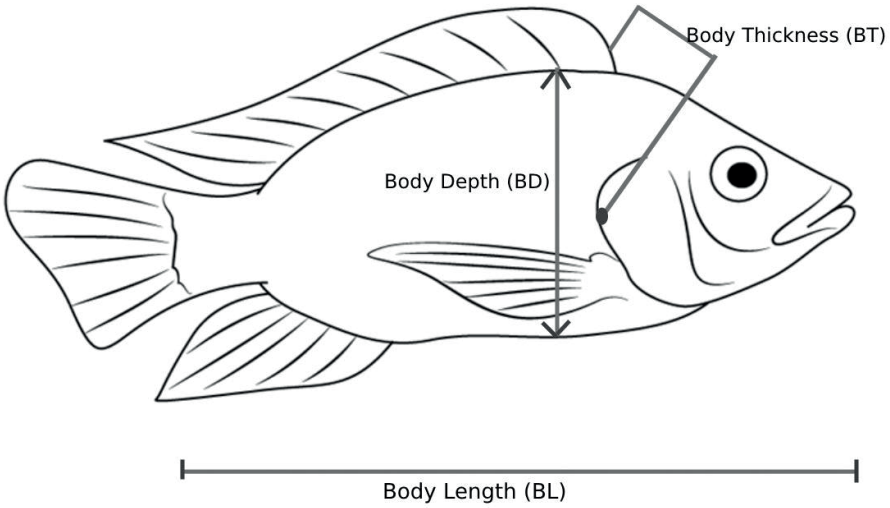


Figure S1.1: Morphometric measurements of GST®. The fish is drawn from www.drawingtutorials101.com

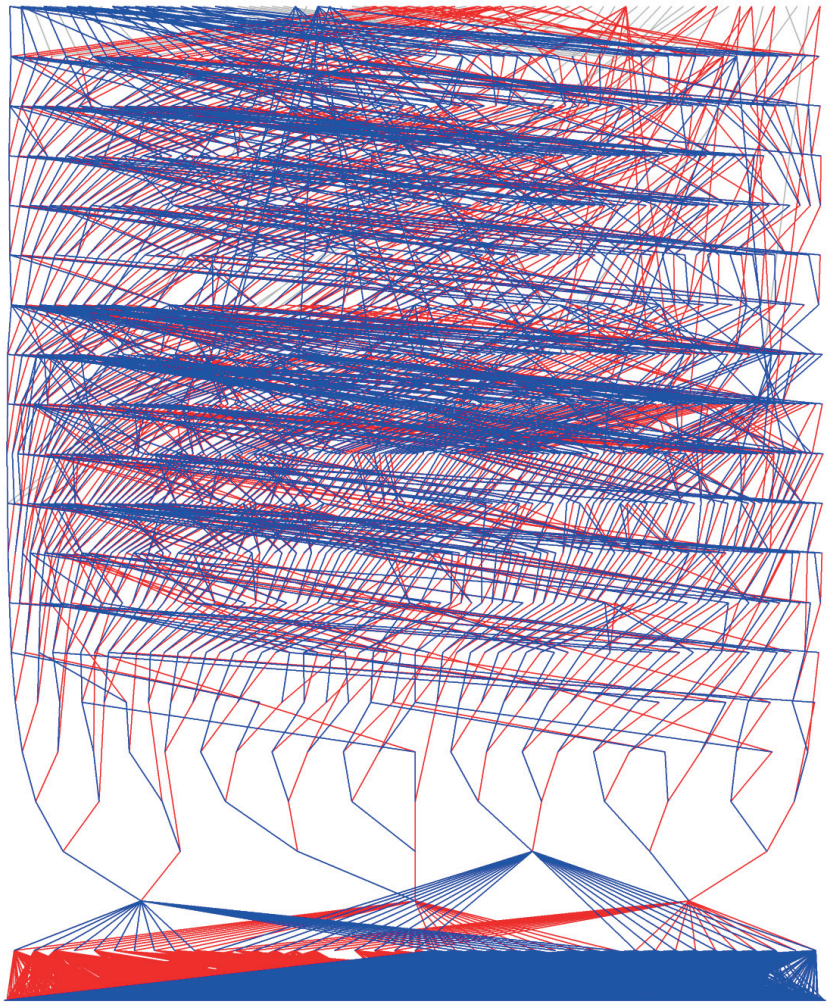


Figure S1.2: Pedigree structure of the 22 generations of GST strain generated by Pedantics package (Morrissey and Wilson, 2010)[1] in R arranged according to the depth of information available. Red, blue and grey lines represent maternal, paternal and uninformative links. The base generation is the Generation 3 of GIFT.

[1] Morrissey MB, Wilson AJ (2010). pedantics: an r package for pedigree-based genetic simulation and pedigree manipulation, characterization and viewing. *Mol Ecol Resour* **10**: 711–719.

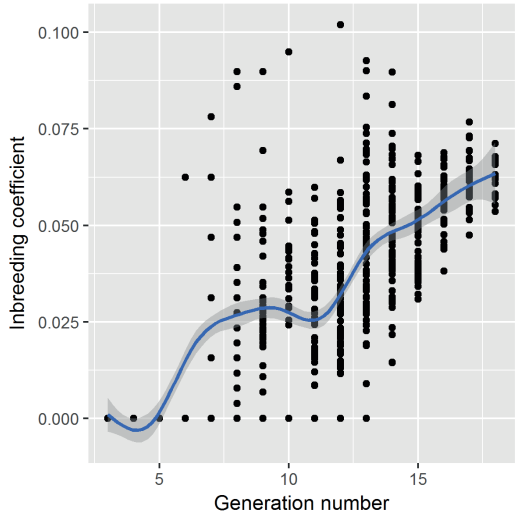


Figure S1.3: Inbreeding coefficients at different generations of pedigree for GST Tilapia. The curve was plotted using loess option in ggplot2 package (Wilkinson, 2011)[2] in R. The shaded area along the blue line represents ± one s.e.

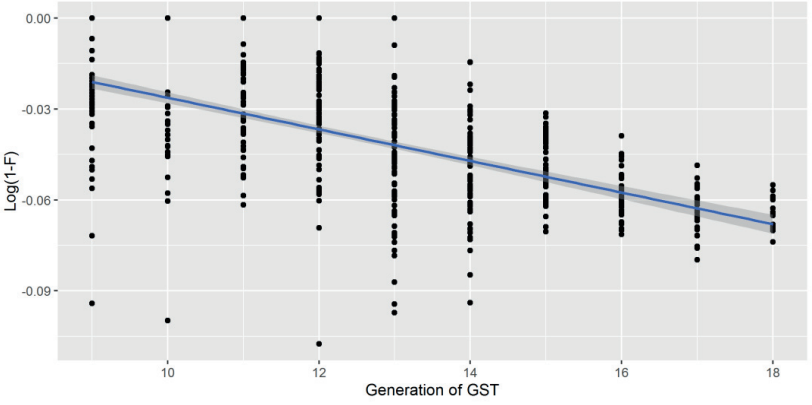


Figure S1.4: Figure showing simple regression coefficient of $\log_e(1\text{-}F)$ over the generation 9 to 18; where F is the inbreeding coefficient. The slope -0.0053 represents $-\Delta F$. N_e of the GST strain was calculated as $1/(2\Delta F) = 95$. The shady area along the blue line shows the standard error of the fitted regression line.

[2] Wilkinson L (2011). ggplot2: Elegant Graphics for Data Analysis by WICKHAM, H. *Biometrics* **67**: 678–679.

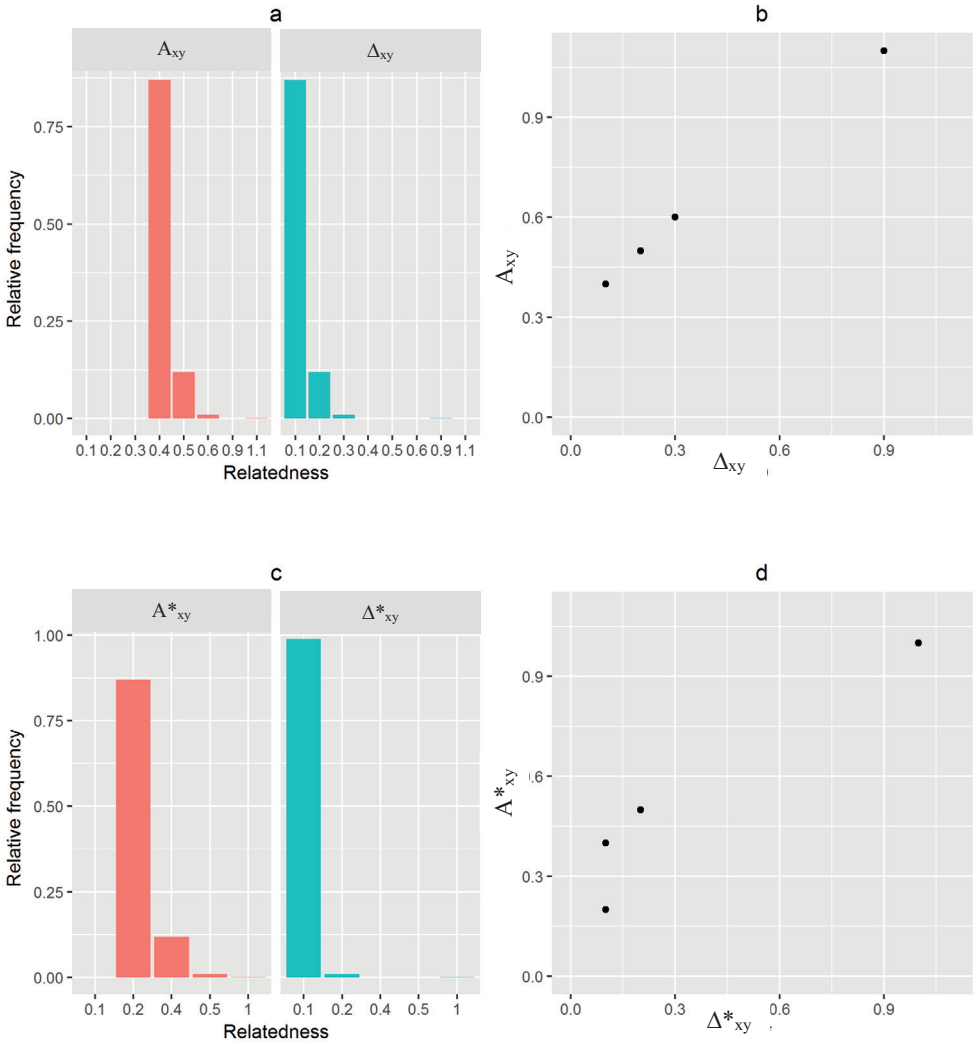


Figure S1.5: Coefficients of fraternity (Δ_{xy}) and numerator relationships (A_{xy}) between animals having phenotypic observations calculated either using the full pedigree (a and b) or only the last 3 generations (c and d) corresponding to a base generation defined by G1 to G4. Plots a and c shows the relative frequency and b and d shows the correspondence between A_{xy} and Δ_{xy}.

Supplement Information 2

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi[1], JA Woolliams[1,2], THE Meuwissen[1] and HM Gjøen[1]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom

Correspondence between SFM, ADM and A*D*M* models.

SFM MODELS

SFM models estimate the 3 core components in the factorial design directly, and these are V_{Sire}, V_{Dam}, and V_{Fsib}. Let C denote cov(y_i, y_j) for individuals *i* and *j*.

i, *j* no common parent (U), $C_U = 0$ (1)

i, *j* paternal half-sibs (PHS), $C_{PHS} = V_{Sire}$ (2)

i, *j* maternal half-sibs (MHS), $C_{MHS} = V_{Dam}$ (3)

i, *j* full-sibs (FS), $C_{FS} = V_{Sire} + V_{Dam} + V_{Fsib}$ (4)

Standard interpretation of these models (Falconer *et al.*, 1996) would imply $\sigma^2_M = C_{MHS} - C_{PHS}$, $\sigma^2_A = 4C_{PHS} = 4V_{Sire}$, and $\sigma^2_D = 4(C_{FS} - C_{PHS} - C_{MHS}) = 4V_{Fsib}$. If components $C_{MHS} < C_{PHS}$ or $C_{FS} < C_{PHS}+C_{MHS}$ further analyses may drop components and pool information, however for the current purpose these initial estimates will be used.

ADM MODELS

The covariance among different type of relationships among the phenotyped individuals relative to a base generation are described by the numerator relationships and coefficients of fraternity relative. Given the (intended) single parents selected from generation 20 these coefficients show no variation (Figure S1.2 in Supplementary 1).

For ADM models with generation 3 base and *i*, *j* in generation 22.

$$C_U = 0.357\,\sigma^2_A + 0.0928\,\sigma^2_D \quad (5)$$

$$C_{PHS} = 0.475\,\sigma^2_A + 0.1610\,\sigma^2_D \quad (6)$$

$$C_{MHS} = 0.475\,\sigma^2_A + 0.1610\,\sigma^2_D + \sigma^2_M \quad (7)$$

$$C_{FS} = 0.592\,\sigma^2_A + 0.2851\,\sigma^2_D + \sigma^2_M \quad (8)$$

Here if *i, j* are unrelated the covariance is the accumulated genotypic drift from the base generation and is accounted for by the mean fitted to the data in the models and equation 1 can be subtracted from the (2), (3) and (4).

$$C_{PHS} = 0.118\,\sigma^2_A + 0.069\,\sigma^2_D \quad (9)$$

$$C_{MHS} = 0.118\,\sigma^2_A + 0.069\,\sigma^2_D + \sigma^2_M \quad (10)$$

$$C_{FS} = 0.235\,\sigma^2_A + 0.193\,\sigma^2_D + \sigma^2_M \quad (11)$$

Now solving equating (9), (10) and (11) and equating them to (2) to (4).

- Estimate of σ^2_M remains unchanged as $C_{MHS} - C_{PHS}$ so moving from SFM to ADM makes no change.
- Estimate of σ^2_D is now $(C_{FS}–C_{PHS}–C_{MHS})/0.055 = 18.18(C_{FS}–C_{PHS}–C_{MHS}) = 18.18V_{Fsib}$ a simple scaling of the result from the SFM model, but 4.56 times that expected from standard assumptions.
- Estimate of σ^2_A is now $19.11C_{PHS}-10.63(C_{FS}-C_{MHS}) = 8.48V_{Sire} -10.63V_{Fsib}$. Therefore the estimate of σ^2_A is unaffected by the estimate of σ^2_M but reduces as the additional variance common to full-sibs increases. As a consequence a positive variance between half-sib families can yield 0 for σ^2_A if $V_{Fsib} > 0.80V_{Sire}$. If $V_{Fsib} = 0$, then σ^2_A is 2.1 - fold greater than predicted from the standard assumptions, primarily due to the sires within parent groups A and B being full-sibs, and the removal of the reciprocal cross effect as a fixed effect.

A consequence of σ^2_M and σ^2_D being estimated from the same quantity as the SMF models is that the statistical significance will be identical, and the s.e.s will have the same scaling factor as the quantity itself.

A*D*M* MODELS

For A*D*M* models with generation 20 base and *i*, *j* in generation 22.

$$C_U = 0.250\,\sigma^2_A + 0.0625\,\sigma^2_D \quad (12)$$

$$C_{PHS} = 0.375\,\sigma^2_A + 0.1250\,\sigma^2_D \quad (13)$$

$$C_{MHS} = 0.375\,\sigma^2_A + 0.1250\,\sigma^2_D + \sigma^2_M \quad (14)$$

$$C_{FS} = 0.500\,\sigma^2_A + 0.2500\,\sigma^2_D + \sigma^2_M \quad (15)$$

Here if *i, j* are unrelated the covariance is the accumulated genotypic drift from the base generation and is accounted for by the mean fitted to the data in the models and equation 1 can be subtracted from the (2), (3) and (4).

$$C_{PHS} = 0.125\,\sigma^2_A + 0.0625\,\sigma^2_D \quad (16)$$

$$C_{MHS} = 0.125\,\sigma^2_A + 0.0625\,\sigma^2_D + \sigma^2_M \quad (17)$$

$$C_{FS} = 0.250\,\sigma^2_A + 0.1875\,\sigma^2_D + \sigma^2_M \quad (18)$$

Now solving equating (16), (17) and (18) and equating them to (2) to (4).

- Estimate of σ^2_M remains unchanged as $C_{MHS} - C_{PHS}$ so moving from SFM to A*D*M* makes no change.
- Estimate of σ^2_D is now $(C_{FS}–C_{PHS}–C_{MHS})/0.0675 = 16(C_{FS}–C_{PHS}–C_{MHS}) = 16V_{Fsib}$ a simple scaling of the result from the SFM model, but 4 times that expected from standard assumptions.
- Estimate of σ^2_A is now $16C_{PHS}-8(C_{FS}-C_{MHS}) = 8(V_{Sire} - V_{Fsib})$, which is similar to the result for ADM models in structure. If $V_{Fsib} = 0$, $\sigma^2_A = 8V_{Sire}$.

Consequently the outcomes of A*D*M* will be similar to those of ADM models but with slightly different scaling factors.

EVIDENCE FROM EVALUATIONS

The EBVs from the full ADM and the full A*D*M* models were plotted for (i) the sires and (ii) all offspring and are shown in Fig. S4.1. It can be seen the EBVs from one model is a linear transformation of the other.

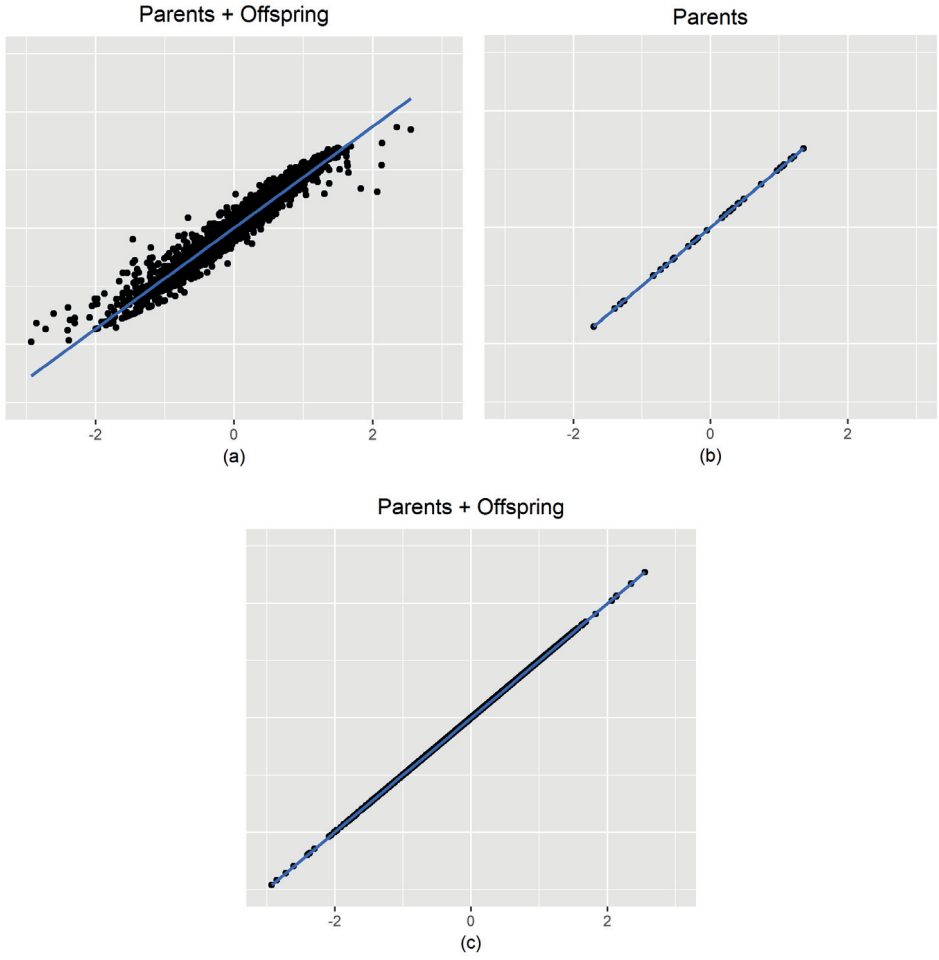


Figure S4.1: Scatterplot for the EBVs obtained from different models. Figure (a) and (b) are the comparison between A model and the model without pedigree information and figure (c) is the comparison between A and A* models.

Supplement Information 3

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi[1], JA Woolliams[1,2], THE Meuwissen[1] and HM Gjøen[1]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom

Supplementary 3: Tables of variance component estimates for all the models

Table S3.1: Heritabilities for different models and different traits (standard errors are inside parenthesis). Values of 0 for component indicate estimate was bound to zero.

Trait	Model	Variance parameters					Ratios		
		Additive	Dominance	Maternal	Residual	Phenotypic	h^2	d^2	m^2
BD	A	0.26 [0.08]	-	-	0.32 [0.04]	0.58 [0.04]	0.44 [0.1]	-	-
	AD	0.15 [0.1]	0.15 [0.09]	-	0.27 [0.05]	0.57 [0.04]	0.27 [0.15]	0.26 [0.15]	-
	ADM	0 [0]	0.14 [0.05]	0.05 [0.02]	0.35 [0.04]	0.53 [0.03]	0 [0]	0.26 [0.09]	0.09 [0.04]
	AM	0.09 [0.04]	-	0.05 [0.02]	0.4 [0.02]	0.54 [0.03]	0.17 [0.07]	-	0.09 [0.04]
	A*	0.24 [0.07]	-	-	0.32 [0.04]	0.56 [0.04]	0.43 [0.1]		
	A*D*	0.16 [0.08]	0.13 [0.08]	-	0.26 [0.05]	0.55 [0.04]	0.29 [0.13]	0.24 [0.14]	
	A*D*M*	0 [0]	0.14 [0.05]	0.05 [0.02]	0.33 [0.04]	0.52 [0.03]	0 [0]	0.27 [0.09]	0.09 [0.04]
	A*M*	0.08 [0.04]	-	0.05 [0.02]	0.40 [0.02]	0.53 [0.03]	0.16 [0.07]	-	0.09 [0.04]
	S	0.12 [0.04]	-	-	0.44 [0.01]	0.47 [0.02]	0.25 [0.07]	-	
	SF	0.11 [0.04]	0.03 [0.02]	-	0.44 [0.01]	0.50 [0.02]	0.23 [0.06]	0.07 [0.04]	
	SFM	0.03 [0.02]	0.04 [0.02]	0.05 [0.02]	0.44 [0.01]	0.51 [0.03]	0.07 [0.04]	0.07 [0.04]	0.09 [0.04]
	SM	0.04 [0.02]	-	0.05 [0.02]	0.44 [0.01]	0.51 [0.03]	0.08 [0.04]	-	0.09 [0.04]

Trait	Model	Variance parameters					Ratios		
		Additive	Dominance	Maternal	Residual	Phenotypic	h^2	d^2	m^2
BL	A	0.97 [0.31]	-	-	2.33 [0.17]	3.29 [0.18]	0.29 [0.08]	-	-
	AD	0.74 [0.43]	0.33 [0.45]	-	2.2 [0.24]	3.27 [0.18]	0.23 [0.12]	0.1 [0.14]	-
	ADM	0.30 [0.4]	0.36 [0.46]	0.12 [0.09]	2.39 [0.22]	3.17 [0.15]	0.09 [0.12]	0.12 [0.14]	0.04 [0.03]
	AM	0.55 [0.26]	-	0.12 [0.09]	2.52 [0.15]	3.19 [0.15]	0.17 [0.08]	-	0.04 [0.03]
	A*	0.90 [0.29]			2.33 [0.17]	3.23 [0.16]	0.28 [0.08]		
	A*D*	0.74 [0.37]	0.28 [0.39]		2.19 [0.26]	3.21 [0.17]	0.23 [0.11]	0.09 [0.12]	
	A*D*M*	0.32 [0.34]	0.32 [0.4]	0.12 [0.09]	2.36 [0.24]	3.12 [0.15]	0.10 [0.11]	0.10 [0.13]	0.04 [0.03]
	A*M*	0.52 [0.25]		0.12 [0.09]	2.52 [0.15]	3.15 [0.14]	0.16 [0.07]		0.04 [0.03]
	S	0.45 [0.15]			2.78 [0.08]	3.00 [0.11]	0.15 [0.05]		
	SF	0.44 [0.15]	0.07 [0.1]		2.77 [0.08]	3.00 [0.11]	0.15 [0.05]	0.02 [0.03]	

	SFM	0.24 [0.12]	0.08 [0.1]	0.12 [0.09]	2.76 [0.08]	3.02 [0.12]	0.08 [0.04]	0.03 [0.03]	0.04 [0.03]
	SM	0.26 [0.12]		0.12 [0.09]	2.78 [0.08]	3.02 [0.12]	0.09 [0.04]		0.04 [0.03]
Trait	**Model**	**Variance parameters**					**Ratios**		
		Additive	**Dominance**	**Maternal**	**Residual**	**Phenotypic**	**h^2**	**d^2**	**m^2**
BT	A	1.97 [0.69]	-	-	7.44 [0.41]	9.41 [0.43]	0.21 [0.07]	-	-
	AD	0.96 [1.13]	1.51 [1.36]	-	6.86 [0.65]	9.34 [0.43]	0.10 [0.12]	0.16 [0.15]	-
	ADM	0 [0]	1.53 [0.7]	0.29 [0.2]	7.3 [0.54]	9.12 [0.35]	0 [0]	0.17 [0.07]	0.03 [0.02]
	AM	1.05 [0.56]	-	0.28 [0.2]	7.87 [0.36]	9.20 [0.38]	0.11 [0.06]	-	0.03 [0.02]
	A*	1.85 [0.65]			7.44 [0.41]	9.29 [0.39]	0.20 [0.06]		
	A*D*	1.08 [0.93]	1.32 [1.19]		6.77 [0.72]	9.17 [0.4]	0.12 [0.1]	0.14 [0.13]	
	A*D*M*	0.07 [0.84]	1.48 [1.22]	0.30 [0.2]	7.15 [0.69]	8.99 [0.36]	0.01 [0.1]	0.16 [0.14]	0.03 [0.02]
	A*M*	0.98 [0.53]		0.28 [0.2]	7.87 [0.36]	9.13 [0.36]	0.11 [0.06]		0.03 [0.02]
	S	0.92 [0.32]			8.36 [0.24]	8.83 [0.28]	0.10 [0.03]		
	SF	0.87 [0.32]	0.33 [0.3]		8.3 [0.24]	8.82 [0.28]	0.10 [0.04]	0.04 [0.03]	
	SFM	0.40 [0.26]	0.37 [0.3]	0.30 [0.2]	8.29 [0.24]	8.88 [0.31]	0.05 [0.03]	0.04 [0.03]	0.03 [0.02]
	SM	0.49 [0.26]		0.28 [0.2]	8.36 [0.24]	8.89 [0.31]	0.06 [0.03]		0.03 [0.02]
Trait	**Model**	**Variance parameters**					**Ratios**		
		Additive	**Dominance**	**Maternal**	**Residual**	**Phenotypic**	**h^2**	**d^2**	**m^2**
BWH	A	3216 [989]	-	-	4312 [496]	7528 [546]	0.43 [0.1]	-	-
	AD	2085 [1219]	1688 [1090]	-	3673 [636]	7447 [543]	0.28 [0.15]	0.23 [0.15]	-
	ADM	0 [0]	1548 [614]	637 [313]	4739 [450]	6925 [381]	0 [0]	0.22 [0.08]	0.09 [0.04]
	AM	1008 [492]	-	635 [316]	5342 [287]	6986 [405]	0.14 [0.07]	-	0.09 [0.04]
	A*	3006 [925]			4320 [495]	7326 [487]	0.41 [0.1]		
	A*D*	2149 [1069]	1475 [953]		3575 [680]	7200 [488]	0.30 [0.13]	0.20 [0.13]	
	A*D*M*	0 [0]	1557 [597]	644 [313]	4579 [494]	6780 [361]	0 [0]	0.23 [0.09]	0.10 [0.04]
	A*M*	943 [460]		636 [316]	5344 [287]	6923 [391]	0.14 [0.06]		0.09 [0.04]
	S	1503 [462]			5822 [165]	6573 [281]	0.23 [0.06]		
	SF	1443 [461]	369 [238]		5756 [165]	6570 [280]	0.22 [0.06]	0.06 [0.04]	
	SFM	378 [222]	402 [243]	746 [315]	5745 [166]	6681 [354]	0.06 [0.03]	0.06 [0.04]	0.10 [0.04]
	SM	4712 [231]		736[316]	5816 [165]	6687 [355]	0.07 [0.03]		0.10 [0.04]

Trait	Model	Variance parameters					Ratios		
		Additive	Dominance	Maternal	Residual	Phenotypic	h^2	d^2	m^2
FW	A	503 [155]	-	-	626 [77]	1129 [85]	0.45 [0.11]	-	-
	AD	425[180]	126 [138]	-	575 [94]	1126 [8]	0.38 [0.14]	0.11 [0.12]	-
	ADM	0 [0]	152 [71]	119[53]	755 [54]	1027 [60]	0 [0]	0.15 [0.07]	0.12 [0.05]
	AM	104 [56]	-	116 [52]	812 [36]	1033 [61]	0.10 [0.05]	-	0.11 [0.05]
	A*	470 [145]			627 [77]	1098 [76]	0.43 [0.1]		
	A*D*	413 [161]	111 [121]		568 [99]	1091 [77]	0.38 [0.12]	0.10 [0.11]	
	A*D*M*	11 [93]	142 [127]	119 [53]	743 [71]	1015 [62]	0.01 [0.09]	0.14 [0.13]	0.12 [0.05]
	A*M*	98 [53]		116 [52]	812 [36]	1026 [61]	0.10 [0.05]		0.11 [0.05]
	S	235 [73]			862 [25]	980 [43]	0.24 [0.07]		
	SF	234 [73]	28 [30]		857 [25]	981 [44]	0.24 [0.07]	0.03 [0.03]	
	SFM	41 [26]	36 [32]	120 [53]	855 [25]	1003 [58]	0.04 [0.03]	0.04 [0.03]	0.12 [0.05]
	SM	49 [26]		116 [53]	861 [24]	1002 [58]	0.05 [0.03]		0.12 [0.05]

Trait	Model	Variance parameters					Ratios		
		Additive	Dominance	Maternal	Residual	Phenotypic	h^2	d^2	m^2
FY	A	2.28 [0.76]	-	-	7.34 [0.44]	9.63 [0.46]	0.24 [0.07]	-	-
	AD	2.28 [0.76]	0 [0]	-	7.34 [0.44]	9.63 [0.46]	0.24 [0.07]	0 [0]	-
	ADM	2.11 [0.94]	0 [0]	0.05 [0.18]	7.42 [0.51]	9.58 [0.48]	0.22 [0.09]	0 [0]	0.01 [0.02]
	AM	2.11 [0.94]	-	0.05 [0.18]	7.42 [0.51]	9.58 [0.48]	0.22 [0.09]	-	0.01 [0.02]
	A*	2.13 [0.71]			7.35 [0.43]	9.48 [0.42]	0.23 [0.07]		
	A*D*	2.13 [0.71]	0 [0]		7.35 [0.43]	9.48 [0.42]	0.23 [0.07]	0 [0]	
	A*D*M*	1.97 [0.88]	0 [0]	0.05 [0.18]	7.43 [0.5]	9.45 [0.43]	0.21 [0.09]	0 [0]	0.01 [0.02]
	A*M*	1.97 [0.88]		0.05 [0.18]	7.43 [0.5]	9.45 [0.43]	0.21 [0.09]		0.01 [0.02]
	S	1.07 [0.36]			8.41 [0.24]	8.95 [0.29]	0.12 [0.04]		
	SF	1.07 [0.36]	0 [0]		8.41 [0.24]	8.95 [0.29]	0.12 [0.04]	0 [0]	
	SFM	0.98 [0.44]	0 [0]	0.05 [0.19]	8.41 [0.24]	8.96 [0.3]	0.11 [0.05]	0 [0]	0.01 [0.02]
	SM	0.98 [0.44]		0.05 [0.19]	8.41 [0.24]	8.96 [0.3]	0.11 [0.05]		0.01 [0.02]

Supplement Information 4

Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits

R Joshi[1], JA Woolliams[1,2], THE Meuwissen[1] and HM Gjøen[1]

[1]Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, 1432 Ås, Norway

[2] The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom

Literature reviews for heritabilities and common maternal and environmental effects in tilapia.

Literature review was done for additive heritability and common maternal and environmental effect for Nile tilapia (GIFT and GST® strains). Most of the literature were found for GIFT strain and wide ranges of values were seen which might be due to different generations of study population or due to the different models used to calculate the values. Nguyen et al. (2010) have reported the common maternal and environmental effect lower than ours for some traits (BD, BWH and FW) in the GIFT strain. Despite the confounding of the maternal and common environmental effects in their model, we are not sure why our values are greater than their values.

Table S3.1: Literature reviews for heritabilities (SE are inside the parenthesis)

Traits	Heritability	Reference	Statement
BD	0.14 (0.037)	(Nguyen *et al.*, 2007)	Female
	0.17 (0.046)	(Nguyen *et al.*, 2007)	Male
	0.20 (0.039)	(Nguyen *et al.*, 2007)	All
	0.28 (0.17–0.41)	(Reis Neto *et al.*, 2014)	
	0.32 (0.10)	(Nguyen *et al.*, 2010)	
BL	0.19 (0.12)	(Rutten *et al.*, 2005)	
	0.29 (0.05)	(Nguyen *et al.*, 2007)	Female
	0.3 (0.19–0.42)	(Reis Neto *et al.*, 2014)	
	0.30 (0.05)	(Nguyen *et al.*, 2007)	All
	0.30 (0.05)	(Nguyen *et al.*, 2007)	Male
	0.31 (0.10)	(Nguyen *et al.*, 2010)	
BT	0.20 (0.08)	(Nguyen *et al.*, 2010)	
	0.25 (0.13)	(Rutten *et al.*, 2005)	
	0.26 (0.047)	(Nguyen *et al.*, 2007)	Female
	0.26 (0.052)	(Nguyen *et al.*, 2007)	Male
	0.29 (0.043)	(Nguyen *et al.*, 2007)	All
	0.29 (0.19 - 0.41)	(Reis Neto *et al.*, 2014)	
BWH	0.06 to 0.48	(Bentsen *et al.*, 2012)	Standard fertilized ponds without feed supplement
	0.17 to 0.44	(Bentsen *et al.*, 2012)	Standard fertilized pond with feed supplement
	0.26 (0.12)	(Rutten *et al.*, 2005)	
	0.31 (0.05)	(Khaw *et al.*, 2016)	log(BWH)
	0.31 (0.11)	(Nguyen *et al.*, 2010)	
	0.33 (0.05)	(Nguyen *et al.*, 2007)	Male
	0.34 (0.07)	(Ponzoni *et al.*, 2005)	
	0.35 (0.05)	(Nguyen *et al.*, 2007)	All
	0.36 (0.05)	(Nguyen *et al.*, 2007)	Female
	0.42 (0.17)	(Bentsen *et al.*, 2012)	Cage culture with feed supplement

	0.68 (0.16)	(Bentsen *et al.*, 2012)	Cage culture with commercial pellet feed
	0.31 (0.12)		
	0.33 (0.02)	(Yalew, 2007)	GST using simple model
FW	0.24 (0.11)	(Rutten *et al.*, 2005)	
	0.33 (0.10)	(Nguyen *et al.*, 2010)	
FY	0.12 (0.06)	(Rutten *et al.*, 2005)	
	0.25 (0.07)	(Nguyen *et al.*, 2010)	
	0.07 (0.03)	(Yalew, 2007)	GST using simple model

Table S3.2: Literature reviews for common maternal and environmental effects

Traits	Common Maternal and environmental effect	Reference	Statement
BD	0.04 (0.04)	(Nguyen *et al.*, 2010)	
	0.15 (0.02)	(Khaw *et al.*, 2012)	Cage
	0.19 (0.019)	(Nguyen *et al.*, 2007)	Female
	0.20 (0.028)	(Khaw *et al.*, 2012)	Pond
	0.24 (0.021)	(Nguyen *et al.*, 2007)	All
	0.26 (0.03)	(Nguyen *et al.*, 2007)	Male
BL	0.05 (0.04)	(Nguyen *et al.*, 2010)	
	0.16 (0.02)	(Nguyen *et al.*, 2007)	All
	0.16 (0.02)	(Nguyen *et al.*, 2007)	Female
	0.16 (0.02)	(Nguyen *et al.*, 2007)	Male
	0.18 (0.03)	(Khaw *et al.*, 2012)	Cage
	0.21 (0.03)	(Khaw *et al.*, 2012)	Pond
BT	0.05 (0.04)	(Nguyen *et al.*, 2010)	
	0.14 (0.02)	(Khaw *et al.*, 2012)	Cage
	0.15 (0.02)	(Nguyen *et al.*, 2007)	Female
	0.16 (0.02)	(Nguyen *et al.*, 2007)	All
	0.17 (0.03)	(Khaw *et al.*, 2012)	Pond
	0.18 (0.02)	(Nguyen *et al.*, 2007)	Male
BWH	0.08 (0.05)	(Nguyen *et al.*, 2010)	
	0.11 (0.02)	(Santos *et al.*, 2011)	By Bayesian method
	0.12 (0.02)	(Khaw *et al.*, 2016)	
	0.15 (0.03)	(Ponzoni *et al.*, 2005)	
	0.18 (0.02)	(Nguyen *et al.*, 2007)	All
	0.18 (0.02)	(Nguyen *et al.*, 2007)	Female
	0.18 (0.03)	(Khaw *et al.*, 2012)	Cage
	0.20 (0.02)	(Nguyen *et al.*, 2007)	Male
	0.26 (0.03)	(Khaw *et al.*, 2012)	Pond
FW	0.05 (0.04)	(Nguyen *et al.*, 2010)	
FY	0.00 (0.00)	(Nguyen *et al.*, 2010)	

References

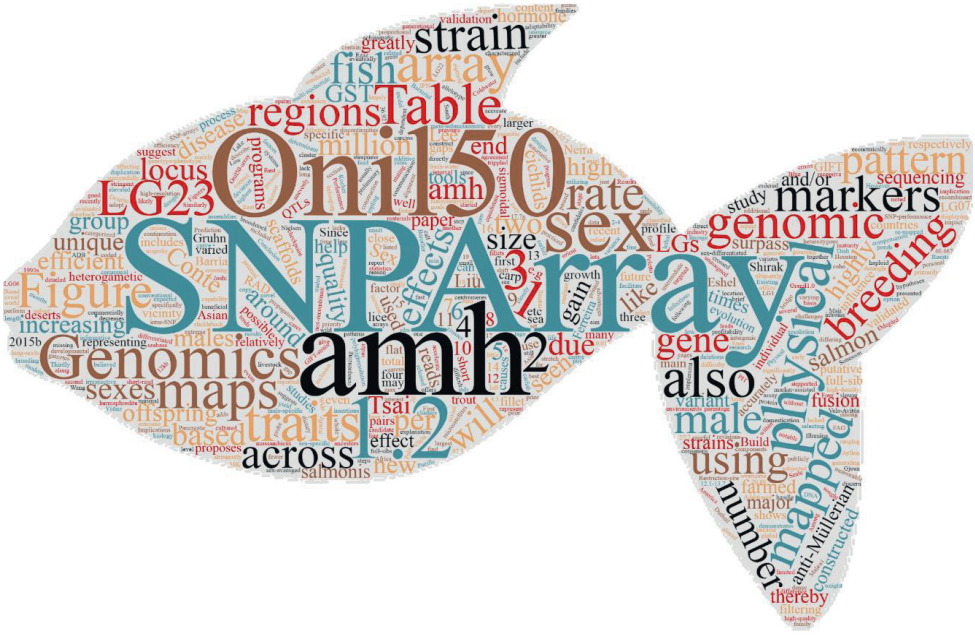
1. Bentsen HB, Gjerde B, Nguyen NH, Rye M, Ponzoni RW, Palada de Vera MS, *et al.* (2012). Genetic improvement of farmed tilapias: Genetic parameters for body weight at harvest in Nile tilapia (Oreochromis niloticus) during five generations of testing in multiple environments. *Aquaculture* **338**: 56–65.
2. Khaw HL, Ponzoni RW, Hamzah A, Abu-Bakar KR, Bijma P (2012). Genotype by production environment interaction in the GIFT strain of Nile tilapia (Oreochromis niloticus). *Aquaculture* **326**: 53–60.
3. Khaw HL, Ponzoni RW, Yee HY, Aziz MA bin, Bijma P (2016). Genetic and non-genetic indirect effects for harvest weight in the GIFT strain of Nile tilapia (Oreochromis niloticus). *Aquaculture* **450**: 154–161.
4. Nguyen NH, Khaw HL, Ponzoni RW, Hamzah A, Kamaruzzaman N (2007). Can sexual dimorphism and body shape be altered in Nile tilapia (Oreochromis niloticus) by genetic means? *Aquaculture* **272**: S38–S46.
5. Nguyen NH, Ponzoni RW, Abu-Bakar KR, Hamzah A, Khaw HL, Yee HY (2010). Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), Oreochromis niloticus. *Aquaculture* **305**: 1–5.
6. Ponzoni RW, Hamzah A, Tan S, Kamaruzzaman N (2005). Genetic parameters and response to selection for live weight in the GIFT strain of Nile tilapia (Oreochromis niloticus). *Aquaculture* **247**: 203–210.
7. Reis Neto RV, Oliveira CAL de, Ribeiro RP, Freitas RTF de, Allaman IB, Oliveira SN de (2014). Genetic parameters and trends of morphometric traits of GIFT tilapia under selection for weight gain. *Sci Agric* **71**: 259–265.
8. Rutten MJM, Bovenhuis H, Komen H (2005). Genetic parameters for fillet traits and body measurements in Nile tilapia (Oreochromis niloticus L.). *Aquaculture* **246**: 125–132.
9. Santos AI, Ribeiro RP, Vargas L, Mora F, Alexandre Filho L, Fornari DC, *et al.* (2011). Bayesian genetic parameters for body weight and survival of Nile tilapia farmed in Brazil. *Pesqui Agropecuária Bras* **46**: 33–43.
10. Yalew DW (2007). Genetic and phenotypic parameter estimation in selected farm tilapia population (Oreochromis niloticus). Norwegian University of Life Sciences (UMB).

3.2Paper II

Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (*O. niloticus*)

Rajesh Joshi, Mariann Árnyasi, Sigbjørn Lien, Hans Magnus Gjøen,
Alejandro Tola Alvarez, Matthew Kent

Frontiers in Genetics | doi: 10.3389/fgene.2018.00472



frontiers in Genetics

ORIGINAL RESEARCH
published: 15 October 2018
doi: 10.3389/fgene.2018.00472

Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile Tilapia (*O. niloticus*)

Rajesh Joshi[1†], Mariann Árnyasi[1†], Sigbjørn Lien[1], Hans Magnus Gjøen[1], Alejandro Tola Alvarez[2] and Matthew Kent[1]*

[1] Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, Ås, Norway, [2] Genomar Genetics AS, Trondheim, Norway

Despite being the second most important aquaculture species in the world accounting for 7.4% of global production in 2015, tilapia aquaculture has lacked genomic tools like SNP-arrays and high-density linkage maps to improve selection accuracy and accelerate genetic progress. In this paper, we describe the development of a genotyping array containing more than 58,000 SNPs for Nile tilapia (*Oreochromis niloticus*). SNPs were identified from whole genome resequencing of 32 individuals from the commercial population of the Genomar strain, and were selected for the SNP-array based on polymorphic information content and physical distribution across the genome using the Orenil1.1 genome assembly as reference sequence. SNP-performance was evaluated by genotyping 4991 individuals, including 689 offspring belonging to 41 full-sib families, which revealed high-quality genotype data for 43,588 SNPs. A preliminary genetic linkage map was constructed using Lepmap2 which in turn was integrated with information from the O_niloticus_UMD1 genome assembly to produce an integrated physical and genetic linkage map comprising 40,186 SNPs distributed across 22 linkage groups (LGs). Around one-third of the LGs showed a different recombination rate between sexes, with the female being greater than the male map by a factor of 1.2 (1632.9 to 1359.6 cM, respectively), with most LGs displaying a sigmoid recombination profile. Finally, the sex-determining locus was mapped to position 40.53 cM on LG23, in the vicinity of the anti-Müllerian hormone (amh) gene. These new resources has the potential to greatly influence and improve the genetic gain when applying genomic selection and surpass the difficulties of efficient selection for invasively measured traits in Nile tilapia.

Keywords: Nile tilapia, linkage map, SNP array, genomics, sex determination, amh, anti-Müllerian hormone

OPEN ACCESS

Edited by:
Ross Houston, University of Edinburgh, United Kingdom

Reviewed by:
José Manuel Yáñez, Universidad de Chile, Chile
Costas S. Tsigenopoulos, Hellenic Centre for Marine Research, Greece

***Correspondence:**
Rajesh Joshi
rajesh.joshi@nmbu.no

†These authors have contributed equally to this work

Specialty section:
This article was submitted to Livestock Genomics, a section of the journal Frontiers in Genetics

Received: *18 May 2018*
Accepted: *24 September 2018*
Published: *15 October 2018*

Citation:
Joshi R, Árnyasi M, Lien S, Gjøen HM, Alvarez AT and Kent M (2018) Development and Validation of 58K SNP-Array and High-Density Linkage Map in Nile Tilapia (O. niloticus). Front. Genet. 9:472. doi: 10.3389/fgene.2018.00472

INTRODUCTION

Nile tilapia (*Oreochromis niloticus*) is an important fresh-water aquaculture species farmed in more than 100 countries including many developing countries in which the species is an essential source of dietary protein (ADB, 2005). Thanks to its fast growth, short generational interval (5 months), relatively small size, adaptability to different environments, and easy to handle, it is also used as a model species for research in fish endocrinology (Seale et al., 2002), physiology (Wright and Land, 1998; Vilela et al., 2003), and evolutionary and developmental biology (Fujimura and Okada, 2007).

Nile tilapia production is supported by more than 20 breeding programs based mainly in South East Asia and some in Africa and America (Neira, 2010). Most of the commercial and farmed Nile tilapia strains are derived from the genetically improved farmed tilapia (GIFT) base strain established in the early 1990s (Eknath et al., 1993). Among these, the Genomar Supreme Tilapia (GST®) strain which has undergone more than 25 generations of selection.

So far Nile tilapia breeding programs have relied on traditional breeding approaches based on easily measurable phenotypes such as weight and length, and have just recently started to implement modern genome-based strategies, such as marker-assisted and genomic selection (personal communication). Compared to livestock species, aquaculture has been slower to adopt genome-based selection tools largely due to a lack of genomic resources such as reference genomes, SNP arrays, and linkage maps. But in species like rainbow trout, salmon, and common carp where genomic selection is being practiced, priority of utilizing genomic information is on selection for disease and parasitic resistance. For example, resistance against Bacterial Coldwater Disease (BCWD) (Vallejo et al., 2015a,b, 2017b), infectious pancreatic necrosis (IPN) (Yoshida et al., 2018) and *Piscirickettsia salmonis* (Yoshida et al., 2017a) in rainbow trout; *Piscirickettsia salmonis* (Bangera et al., 2017) and resistance against sea lice (Ødegård et al., 2014; Tsai et al., 2016; Correa et al., 2017a) in Atlantic salmon; *Piscirickettsia salmonis* in coho salmon (Barría et al., 2018); and juvenile growth rate in common carp (Tsai et al., 2015b; Palaiokostas et al., 2018). Similarly, lots of GWAS studies have been conducted in these species, primarily for disease resistance (Correa et al., 2015; Liu et al., 2015; Palti et al., 2015; Vallejo et al., 2017a; Barría et al., 2018), resistance against sea lice (Davidson and Yáñez, 2016; Correa et al., 2017b), sexual maturity (Ayllon et al., 2015; Gutierrez et al., 2015) and some carcass quality traits (Sodeland et al., 2013; Tsai et al., 2015a; Gonzalez-Pena et al., 2016; Yoshida et al., 2017b).

The first genome assembly for *O. niloticus* (released in 2011; Orenil1.0, and updated to Orenil1.1 at the end of 2012 (NCBI, 2018)) was based on short-read sequencing. A newer assembly (O_niloticus_UMD1) was generated using a combination of novel long-reads (generated using Pacific Biosciences Technology) and publicly available Illumina short reads (Conte et al., 2017). Four linkage maps of varying resolution were constructed using markers found with Restriction-site Associated DNA (RAD) sequencing (Palaiokostas et al., 2013), microsatellites and/or AFLP markers (Kocher et al., 1998; Lee et al., 2005; Guyon et al., 2012). The RAD based strategies usually generate a SNP resource of medium density and are highly efficient in species where a reference genome is not available (Robledo et al., 2017). In comparison, a SNP-array offers the advantages of increased genotype accuracy of much higher numbers of markers as well as control over the physical distribution of these across the genome (Robledo et al., 2017). In this paper, we report the development of a 58K SNP-array (Onil50) and construction of a high density linkage map in the commercial strain of Nile tilapia, Genomar Supreme Tilapia (GST®), which is the continuation of the widespread GIFT-strain.

MATERIALS AND METHODS

SNP-Array (Onil50) Development

Origin of Sequenced Fish

The GST® strain of Nile tilapia used in this study originates from the original GIFT population (Eknath et al., 1993). This strain was selected for growth from generation 1 to 14, growth and filet yield from generation 15 to 19, and growth, yield, and robustness from generation 20. Thirty-two individuals (13 males and 19 females) from this population were selected for whole genome sequencing. Twenty of them are from generation 23, selected at random from the breeding nucleus and the rest 12 are from a commercial line formed from generation 20 and selected for growth (**Supplementary Figure 6**). The graphical summary of the methodology is given in **Figure 1**.

Whole Genome Sequencing and SNP-Detection

Genomic DNA from these 32 individuals was extracted from fin-clips (preserved in Ethanol) using Qiagen DNeasy columns (Qiagen, Germany). DNA quality was assessed by agarose gel electrophoresis and quantified using a Qubit fluorometer (Thermo Fisher Scientific, United States). After normalization, sequencing libraries were prepared and barcoded using TruSeq sample preparation kit and sequenced (2 × 125) across 10 lanes on an Illumina HiSeq 2500 (Illumina, United States) by a commercial provider. At the time this work was carried out, Orenil1.1 Tilapia represented the highest quality reference genome available (NCBI Assembly *Oreochromis niloticus*: GCF_000188235.2_Orenil1.1_genomic), and reads were aligned to it using BWA-MEM algorithm in BWA 0.7.12 (Li, 2013) with default parameters. Putative SNPs were identified using FreeBayes v0.9.20 (Garrison and Marth, 2012) with parameters genotype-qualities and experimental-gls. Using vcffilter, SNPs with a QUAL score value (phred) of $\leq$20 were removed.

SNP-Filtering

The initial set of putative SNPs was divided into three groups including SNPs located on scaffolds assigned to linkage groups (LGs) of the assembly, SNPs on unassigned scaffolds, and SNPs detected within the mitochondrial genome. SAMtools v1.2/bcftools (Li et al., 2009) was then used to filter out variants according to the following criteria: a SNP was removed if; (i) located within 5 bp to an indel, (ii) had more than one alternative allele, (iii) the sequencing depth exceeded 700 reads, (iv) its alleles were A and T, or C and G (these require twice as many 'probes' on Affymetrix SNP arrays as other SNP allele combinations), (v) if sample genotype quality (GQ) was <30, (vi) minor allele frequency (MAF) <0.05, (vii) all samples were heterozygous for the given SNP, (viii) the variant was detected in fewer than 28 of the samples sequenced. Finally, Hardy–Weinberg Equilibrium (HWE) exact test was calculated using PLINK2 (Chang et al., 2015) and SNPs that showed departure from HWE ($P < 0.05$) were removed.

SNP-Selection

After filtering, 2.76 million putative high-quality SNPs remained. Based on their relationship to genes and physical distribution,

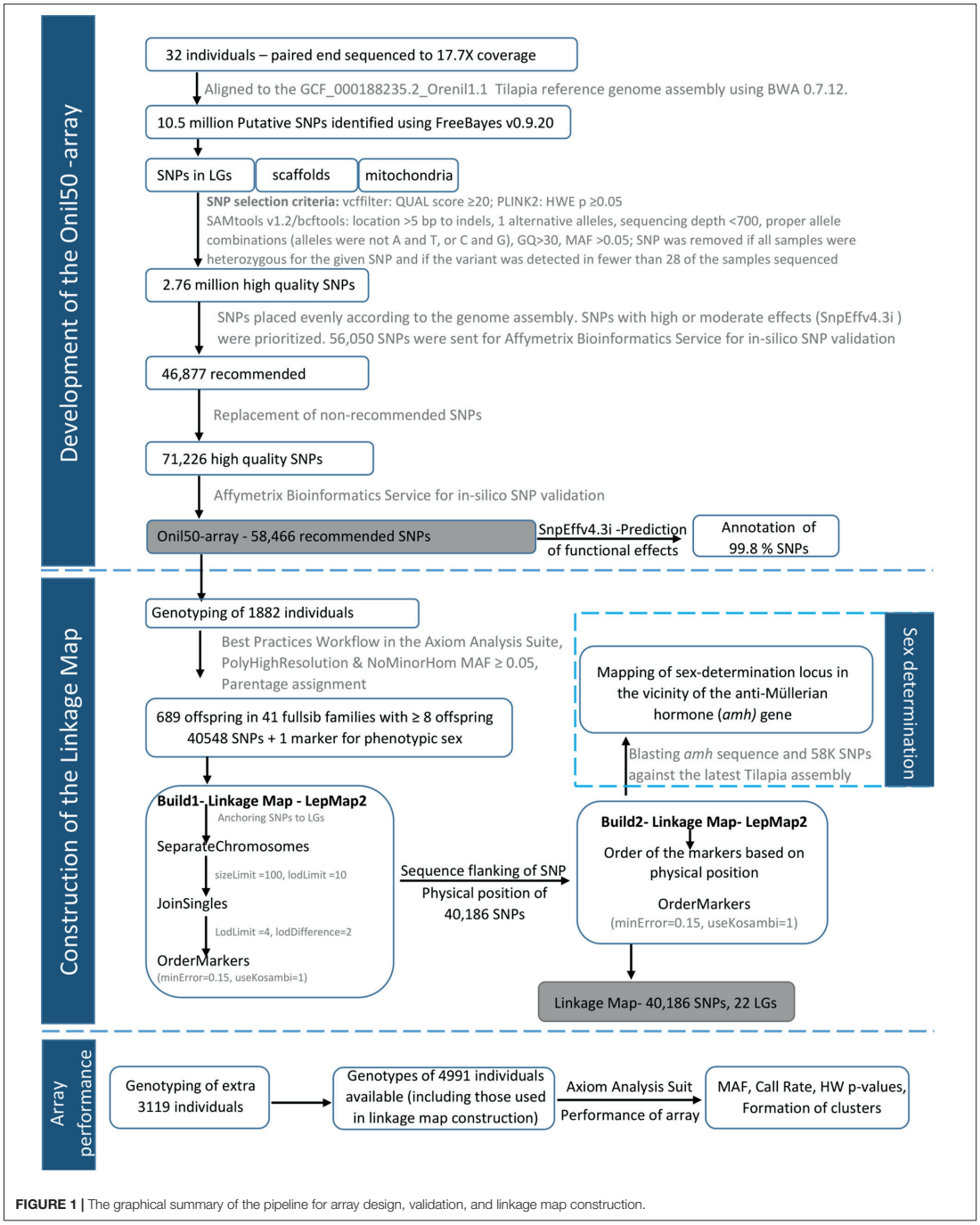


FIGURE 1 | The graphical summary of the pipeline for array design, validation, and linkage map construction.

a subset of these was identified for inclusion on the array. SNPEff v 4v1l (Cingolani et al., 2012) was used to identify SNPs with high and moderate effects (including for example non-synonymous variants). The SNPs with high effects are assumed to have disruptive impact in the protein codification, probably causing protein truncation, loss of function or triggering non-sense mediated decay, e.g., stop gained, frameshift variant, etc.; whereas the SNPs with moderate effects are assumed to be non-disruptive, but they might change protein effectiveness, e.g., missense variant, inframe deletion, etc. From the list of almost 38,000 variants with high and moderate effects, approximately 10,000 were chosen avoiding SNPs within 10 kb of another. A python script was used to fill in gaps and produce a relatively even distribution of SNPs selected at ≈12 kb intervals across the 22 LGs, and ≈33 kb across unmapped scaffolds >50 kb in length. The script was designed to fill a distribution gap with a variant falling within a small size selection window with highest MAF being the main criteria. SNPs from the mitochondrial genome were selected manually. The selected subset of SNPs ($n = 56{,}050$) were submitted to *in silico* validation for Affymetrix Bioinformatic Service. The Affymetrix *in silico* probe set design and evaluation pipeline predicts the performance of SNPs and calculates a conversion probability value (*p*-convert value: representing the probability of a given SNP converting to a reliable SNP assay on the Axiom array system) using various criteria including: binding energy, GC content, and the expected degree of non-specific hybridization to multiple genomic regions. Based on the *p*-convert values, they classify the SNPs into different categories: recommended, neutral, not-recommended, etc. 46,877 SNPs under the categories recommended and/or neutral from probe scoring recommendations were retained and 24,349 extra SNPs were chosen from the regions were the SNPs were discarded. A total of 71,226 SNPs (46,877 + 24,349) were sent back to Affymetrix for *in silico* SNP validation. Finally, the best 58,466 SNPs were chosen to tile on the array based on their probe scoring recommendation (at least one of the strand were recommended, or got neutral category). Upon its release, SNP positions were redefined based on the O_niloticus_UMD1 assembly (Conte et al., 2017) using NCBI's Genome Remapping Service.

Construction of Genetic Map

Genotyping

Genomic DNA was isolated from ethanol-preserved fin clips collected from 1882 Nile tilapia samples using Qiagen 96 DNeasy Blood & Tissue Kits according to manufacturer's instructions (Qiagen, Germany) for map construction. These samples were from different generations of GST® strain within the same breeding population. After quantification and quality checking of DNA, samples were genotyped on the Onil50 array at Center for Integrative Genetics (CIGENE) in Norway.

The unfiltered dataset contained 58,466 SNPs, which were analyzed using the Best Practices Workflow with default settings (sample Dish QC ≥ 0.82, QC call rate ≥97; SNP call-rate cutoff ≥97) in the Axiom Analysis Suite software. Ten samples were excluded from analyzed dataset because of the low call rate. Then, the SNPs classified as PolyHighResolution or NoMinorHom [most informative categories from Best Practices Workflow in Axiom Analysis Suite software (Thermo Fisher Scientific Inc, 2018)] were selected, leaving us with 43,014 SNPs.

Family Structure

The 1872 genotyped individuals could be divided into two groups based on the generations of the breeding population. Group 1 ($n = 1124$) comprised individuals collected following the branching of the 20th generation, and were factorially crossed against each other after two generations. The experimental design for Group 1 is described in Joshi et al. (2018) and was primarily intended to partition the non-additive genetic effects in this population. Fish from Group 2 ($n = 748$) were obtained from the 24th and 25th generations of GST® (**Supplementary Figure 6**).

Parentage assignment was done using an exclusion method which eliminates animals from a list of potential parents when there are opposing homozygotes between parents and offspring (Hayes, 2011). We used all the 43,014 SNPs and permitted a maximum of 100 conflicts between parents and offspring, representing approximately 0.24% of all genotypes. A total of 689 offspring was divided among 41 full-sib families containing ≥8 offspring (mean offspring per family, $\mu = 16.81$). Group 1 (468 offspring with 19 parents) had 34 full-sib families ($\mu = 13.77 \pm 5.5$) and Group 2 (221 offspring with 14 parents) had 7 full-sib families ($\mu = 31.57 \pm 7.23$). The structure of Groups 1 and 2 is shown in **Supplementary Tables 1**, **2**.

Linkage Map Construction

The SNPs displaying a MAF ≤0.05 (2,466 SNPs) were further filtered for linkage map construction. All the retained SNPs ($n = 40{,}548$) had SNP call rate >0.97, so this criteria was not used for filtration. Phenotypic sex was known for a subset of families (221 offspring + 33 parents) and was coded as 12 for males and 11 for females and included in the genotype file ($n = 40{,}549$) before running Lepmap2 (Rastas et al., 2013) for linkage map construction. Lepmap2 uses information from full-sibs and their parents to assign SNP markers to LGs, and applies standard hidden Markov model (HMM) to compute the likelihood of the marker order within each LGs. First, the SNPs were used to construct the preliminary linkage map (Build 1), which was used to anchor, order, and orient the scaffolds in the O_niloticus_UMD1 assembly and upgrading this assembly to O_niloticus_UMD_NMBU (Conte et al., 2018). Eventually, the final physical integrated genetic linkage map (Build 2) was constructed from the order of the markers based on the physical position of the SNPs in O_niloticus_UMD_NMBU assembly.

Build 1: To Anchor SNPs to Different LGs

SeparateChromosomes (a module in Lepmap2) was run testing lodLimits from 1 to 50 and a sizeLimit = 100; a lodLimit of 10 resulted in 22 LGs, also with the lowest number of markers not assigned to any LG. JoinSingles was run to assign single markers to LG groups and tested with lodLimits from 1 to 15 and lodDifference = 2; a lodLimit of 4 was selected as this joined the highest number of single markers. OrderMarkers was used to order the markers within each LG. Each LG

was ordered separately and replicated 5 times with commands: numThreads = 10, polishWindow = 30, filterWindow = 10, useKosambi = 1, minError = 0.15, and the order with highest likelihood was selected as the best order. For sex averaged map OrderMarkers was run similarly by adding sexAverage = 1.

Build 2: Integrated Linkage Map Based on the Order of the SNPs in the New Assembly

Sequence containing each SNP was used to find the physical position of the SNPs in the O_niloticus_UMD_NMBU assembly. Physical position information was used to adjust the order of the SNPs within respective LGs and Lepmap2 was rerun to produce the final linkage map.

Array and SNP-Performance

To get a more comprehensive overview about the array and SNP performance, 3119 additional Nile tilapia samples were genotyped using Onil50 array. The raw dataset of the 1872 samples which were used for linkage mapping was combined with the dataset of the 3119 samples and were analyzed together using the Best Practices Workflow with default parameters in Axiom Analysis Suite software (Thermo Fisher Scientific Inc, 2018). Four quality parameters were assessed on those samples filtered through Dish QC (DQC $\geq$ 0.82), QC call rate (QC CR $\geq$ 97) and plate QC (Percent of passing samples $\geq$50 and average call rate for passing samples $\geq$50) criteria: MAF, SNP call rates, Hardy Weinberg (HW) *p*-values, and clustering. SNPs could be divided into six different types on the basis of formation of clusters (i) "PolyHighResolution" – formation of three clusters with good resolution; (ii) "NoMinorHom" – formation of two clusters with no samples of one homozygous genotype; (iii) "MonoHighResolution" – a single cluster of a homozygous genotype; (iv) "OTV," off-target variants – three good clusters, with a single additional off-target cluster caused by variants in the SNP flanking region; (v) "CallRateBelowThreshold" – the SNP call rate was below the threshold (0.970), but other cluster properties were above the threshold; and (vi) "Other" – the SNPs were not grouped into any of the previous categories.

SnpEffv4.3i (Cingolani et al., 2012) was used to predict functional effects of the 58,340 SNPs which were remapped to O_niloticus_UMD1 assembly.

RESULTS

SNP Selection and Array Development

The sequencing of 32 Nile tilapia generated 4.22×10^9 reads representing an average of 17.7× coverage per individual (stdv = 4.2, min = 9.4, and max = 27.7). After alignment, on average 98% of reads were mapped to the Orenil1.1 assembly yielding 12.78 million variants of which 10.5 million were putative SNPs. Rest 2.2 million were insertions, deletions, multi-nucleotide polymorphisms, etc. After performing multiple steps of filtering described in the section "Materials and Methods," a subset of 2.76 million SNPs was retained and a final set of 58,466 SNPs were selected for assay design and printed on the Onil50 array.

Around 99.8% of the SNPs from the array were successfully re-mapped to the new O_niloticus_UMD1 assembly (**Table 1**). Remapping revealed an increase in the number of SNPs mapping to LGs and a corresponding decrease in the number of SNPs on unmapped scaffolds. The average variant density per LG on the Orenil1.1 assembly is 12.5 $\pm$ 0.35 kb (12.1–13.7 kb). However, since the O_niloticus_UMD1 assembly includes an additional 87 Mb assigned to LGs the average variant density increased to 15.5 $\pm$ 4.06 kb (13.3–32.3 kb) (**Table 1**). Additional information about inter-SNP distance and standard deviation can be found in **Supplementary Table 4**. Physical size of LG03 increased by 2.4 times in the new assembly, thereby increasing the number of SNPs assigned to this LG by 2.3 times.

Performance and Validation of the SNPs in the Array

A total of 4947 samples out of 4991 passed the Dish QC threshold. A total of 4858 samples (97.3%) were left after being subjected to sample call rate. Based on the cluster profile, over 74% of the 58,466 SNPs were classified as PolyHighResolution. More detailed information about the sample and SNP statistics are shown in **Figure 2**.

Prediction of functional effects of the 58,340 remapped SNPs from the Onil50 array resulted, in most cases, in multiple annotations per variant. The effects with the highest putative impact are included for summary in **Table 2**. The majority of the SNPs are intronic (36.92%) or intergenic (20.80%) variants, while about 15% of are non-synonymous mutations. Since the SNP selection process specifically targeted variants with a potential functional effect, these variants are expected to have direct effect on traits of interest.

Linkage Map

A total of 40,549 SNPs were retained following quality filtering, and 99.78% of these (n = 40,467) were ordered within the 22 LGs corresponding to the karyotype of Nile tilapia (**Supplementary Figure 1**). Since, Build 1 linkage map is an intermediate step for the extension of the O_niloticus_UMD1 to the O_niloticus_UMD_NMBU genome assembly (Conte et al., 2018), which is not the aim of this paper, we give only a brief summary of the results. The genetic and physical maps were generally in good agreement with a correlation of $\geq$0.96 between the reference genome position and the genetic map position of the SNPs (**Supplementary Figure 1**). This high correlation with the physical map demonstrates that the genetic map is of high quality and is highly accurate.

A total of 40,186 SNPs mapped to 22 LGs in Build 2 linkage map. The consensus (sex-averaged) map adds up to 1469.69 cM, with individual LG lengths ranging from 56.04 cM (LG19) to 96.68 cM (LG07) (**Table 3**). The average genetic distance across the LGs was 66.8 cM. The number of markers per LG varied from 1349 to 3391, with an average of 1827 markers per LG (**Table 3**). As a consequence of the SNP selection, which sought to position a SNP every 12 kb, the number of markers was mostly proportional to the size of the LG (**Figure 3**). A notable exception is LG03 where the inclusion of previously unassigned

TABLE 1 | Sequence similarity based assignment of SNPs contained on the Onil50-array to Orenil1.1 and O_niloticus_UMD1 genome assemblies.

Orenil1.1 assembly				O_niloticus_UMD1 assembly			
LG	**Length (bp)**	**Variants**	**Variants rate (bp/variant)**	**LG**	**Length (bp)**	**Variants**	**Variants rate (bp/variant)**
LG01	31194787	2571	12133	LG01	38372991	2830	13559
LG02	25048291	2043	12261	LG02	35256741	2395	14721
LG03	19325363	1415	13658	LG03	68550753	2105	32565
LG04	28679955	2288	12535	LG04	38038224	2427	15673
LG05	37389089	2927	12774	LG05	34628617	2549	13585
LG06	36725243	2891	12703	LG06	44571662	2932	15202
LG07	51042256	4128	12365	LG07	62059223	4682	13255
LG08–24	29447820	2314	12726	LG08	30802437	2307	13352
LG09	20956653	1732	12100	LG09	27519051	1909	14415
LG10	17092887	1414	12088	LG10	32426571	1878	17267
LG11	33447472	2653	12607	LG11	36466354	2662	13699
LG12	34679706	2753	12597	LG12	41232431	2833	14554
LG13	32787261	2647	12387	LG13	32337344	2275	14214
LG14	34191023	2700	12663	LG14	39264731	2679	14656
LG15	26684556	2180	12241	LG15	36154882	2255	16033
LG16–21	34890008	2777	12564	LG16	43860769	2848	15401
LG17	31749960	2609	12169	LG17	40919683	2873	14243
LG18	26198306	2075	12626	LG18	37007722	2307	16041
LG19	27159252	2223	12217	LG19	31245232	2301	13579
LG20	31470686	2491	12634	LG20	36767035	2635	13953
LG22	26410405	2083	12679	LG22	37011614	2272	16290
LG23	20779993	1603	12963	LG23	44097196	2225	19819
Total	657350972	52517	12517	Total	868591263	56216	15451
Unmapped scaffold (n = 557)	246010115	5939	41422	Unmapped scaffolds (n = 284)		2151	
Mitochondrial genome	16627	10	1662	Mitochondrial genome	16627	10	
Total number of SNPs on the array		58466		Remapped SNPs in total		58340	
				SNPs failed to remap		126	

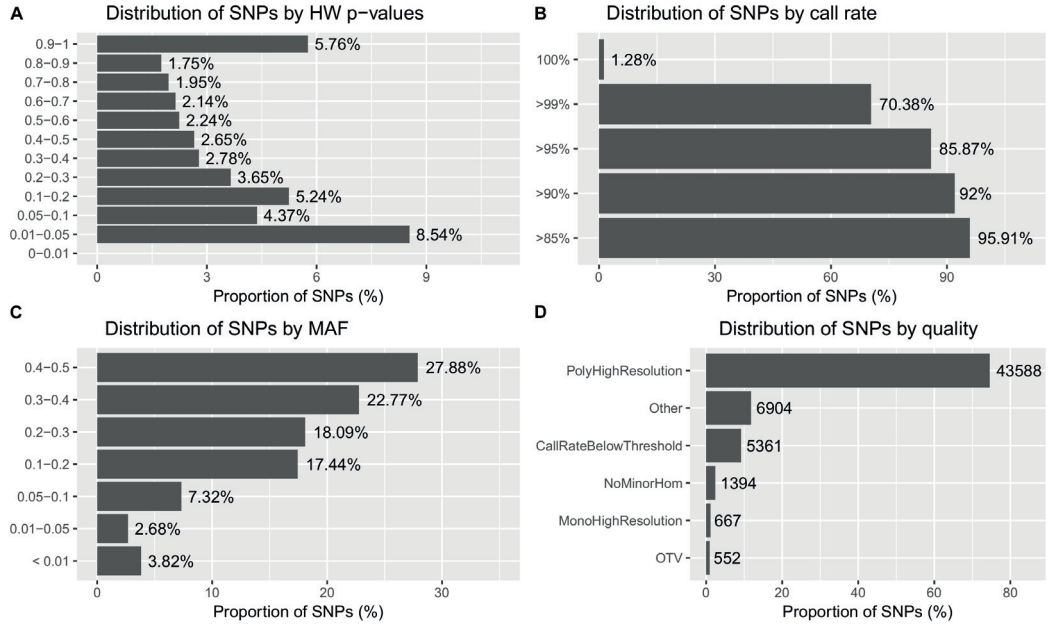


FIGURE 2 | Summary of the SNP metrics based on 58,466 SNPs on 4,947 GST® samples after being subjected to sample call rate in Best Practices Workflow of Axiom Analysis Suite software (see methodology). **(A)** Distribution of SNPs based on HW *p*-values obtained from the exact HWE test. **(B)** Distribution of SNPs based on SNP call rate. **(C)** Distribution of SNPs based on minor allele frequency (MAF). **(D)** Distribution of SNPs based on types of cluster formation (quality). The different types of clusters are based on the quality of the SNPs and has been described in the methodology.

scaffolds has trippled the physical size without a corresponding tripling of SNP numbers. The SNP density (SNPs/Mb) varied across the genome from 19.68 to 56.83 (see also **Figure 4** and **Supplementary Figures 2–4**).

In this study, paternal and maternal informative markers were used to construct specific male and female maps (**Table 3**). Around one-third of the LGs showed a different recombination rate between sexes (**Supplementary Figure 5**), with male and female map lengths differing by a factor of 1.2 (1359.6 and 1632.9 cM, respectively). Generally, female maps were found to be larger in all LGs with the exception of LG02, LG06, and LG22. Sigmoidal pattern of recombination, with no recombination at both ends of the LGs, was seen in almost all LGs (**Figure 4**).

DISCUSSION

High-Density Linkage Map for Nile Tilapia

Existing linkage maps for Nile tilapia contain relatively few markers unevenly distributed across LGs (**Supplementary Table 3**). As a consequence, regions in the genome have poor SNP coverage. By stringently selecting SNPs with an even physical distribution in the genome the linkage map presented includes 10 times more SNPs and fewer gaps, compared to the most recent map (Palaiokostas et al., 2013).

Ferreira et al. (2010) categorized the karyotypes of *O. niloticus* into 3 meta-submetacentric and 19 subtelo-acrocentric chromosomes. The steepness of the curve in **Figure 4** shows the recombination level, with flat lines representing little or no recombination, which may suggest the possible location of the centromeres. There are some discontinuities present in LG03 and LG15, suggesting that our linkage map lacks the SNPs in those regions. These gaps might be due to missing sequence in the assembly or due to the assembly errors. Wide recombination deserts (areas with no recombination) are seen in the initial and/or end regions of most of the LGs, generally up to 5 Mb and sometimes up to 10 Mb (e.g., LG09 and LG10), indicating the presence of subtelo-acrocentric chromosomes. Because of these recombination deserts, most of the LGs, irrespective of the sexes, showed sigmoidal pattern, which is unusual when compared to other fish species. In channel catfish (Li et al., 2014), salmon (Tsai et al., 2015b), Asian seabass (Wang et al., 2015) and stickleback (Roesti et al., 2013) the recombination rates were generally elevated toward the end of the LGs. The possible explanation might be that the GST® strain used in this study is derived from the GIFT strain, formed from crossing among four wild and four cultured Asian strains (Eknath et al., 1993). When the individuals from two different strains are crossed together, the offspring is

TABLE 2 | Summary of annotation for SNPs in the Onil50-array.

SNP categories	Count	Percent
Total number of SNPs in the array	58,446	
Annotation possible	58,340	99.82
Annotation results		
Nonsense-mediated decay (NMD)	19	0.03
Loss of function (LOF)	114	0.20
Intergenic region	12,156	20.80
Intragenic variant	126	0.22
Intron variant	21,581	36.92
Non-synonymous variant		
Missense variant	8,765	15.00
Missense variant & splice region variant	263	0.45
Stop gained	16	0.03
Stop lost	13	0.02
Synonymous variant	1,142	1.95
Non-coding transcript exon variant	27	0.05
Splice acceptor variant & intron variant	9	0.02
Splice donor variant & intron variant	13	0.02
Splice region variant	8	0.01
Splice region variant & intron variant	163	0.28
Splice region variant & non-coding transcript exon variant	7	0.01
Splice region variant & synonymous variant	28	0.05
Upstream gene variant	9,231	15.79
3 prime UTR variant	1,533	2.62
5 prime UTR premature start codon gain variant	21	0.04
5 prime UTR variant	459	0.79
Downstream gene variant	2,646	4.53

heterozygous and causes difficulty in recombination producing stretch with low recombination, which might have resulted to the unique recombination pattern. Low recombination at the end of LGs were also observed in the hybrid crosses of Lake Malawi cichlids (Conte et al., 2018).

Nile tilapia was shown to have a sex-specific pattern of recombination with the female map generally being larger than the male map (Lee et al., 2004). The genetic basis for the differences in the recombination rate between sexes has still not been found, but Li et al. (2014) has listed three major hypotheses. First, the selection perspective hypothesis (Lenormand and Dutheil, 2005; Gruhn et al., 2013), proposes that the selection pressure is higher in male compared to female gametes during the haploid life stage and this male-specific selection leads to decrease in the male recombination rate to maintain the beneficial haplotypes. Second, the compensation hypothesis (Coop and Przeworski, 2007), proposes that the recombination rate is higher in females compared to males to compensate for the less stringent checkpoint for the non-recombinant (achiasmatic) chromosomes. Third, the recombination pathway hypothesis (Gruhn et al., 2013), suggests that the chromatin differences established prior to the onset of the recombination pathway causes the differences in the recombination between the two sexes.

LG23 showed a unique recombination pattern, a flat line of around 5 Mb, in the center of the LG, for which there also is a sex difference in recombination rate. In *O. niloticus*, major XY sex determining regions have earlier been mapped to LG1 (Palaiokostas et al., 2013) and LG23 (Karayücel et al., 2004; Shirak et al., 2006; Eshel et al., 2011, 2012). Further, tandem duplication of the variants of the gene anti-Müllerian hormone (*amh)* in LG23 has been identified as the male sex determinant in Nile tilapia (Li et al., 2015). These variants of *amh* gene have been mapped to around 35.4 Mb region of Nile tilapia genome (discussed below in section "Sex Locus Mapped in the Vicinity of *amh* Gene"), which is the same region where the unique recombination pattern is seen, suggesting limited recombination around the sex-determining genes in *O. niloticus*. Further, LG23 was formed by the fusion of two LGs during the evolution of cichlids (Liu et al., 2013), which might be another reason for this unique recombination pattern.

The fusion of the LGs during the evolutionary process also has an effect on the size of the LGs, as it is believed that the ancestors of cichlids had 24 chromosome pairs, which eventually became 22 pairs (Majumdar and McAndrew, 1986; Ferreira et al., 2010). Our genetic map shows that LG07 is the largest, which has been shown to be formed by the fusion of two LGs during lineage evolution (Poletto et al., 2010; Liu et al., 2013; Conte et al., 2017).

Array Content and Performance

SNP performance was validated by genotyping around 5000 individuals from different generations of the GST® strain of Nile tilapia. Around 75% of the SNPs on the array perform well generating three highly differentiated allelotype clusters (i.e., PolyHighResolution). Around 9% of the SNPs were found to depart from HWE ($p < 0.01$), but it has to be noted that the population genotyped for the validation is the commercial strain that has undergone up to 25 generations of selection. Hence, these departures might be important as they could represent regions under selection, domestication and the outcome of assortative mating (Gutierrez et al., 2016; Adenyo et al., 2017). Whereas the extreme departures might suggest lethal recessive mutations and/or recent mutations or copy number variants (Lee et al., 2008; Graffelman et al., 2017).

For future revisions, the array could be improved by increasing the SNP density in highly recombinant regions of specific LGs like including LG03 and LG23. The use of genetic distance rather than the physical distance to select the SNPs is probably the best option for equidistant SNP distribution across the genome.

Sex Locus Mapped in the Vicinity of *amh* Gene

Sex determination is one of the important aspect in commercial tilapia production, as males are found to grow faster than females and unisex production is a main method to avoid propagation in production ponds or cages. Sex determination in fish is more complicated than mammals as it tends to be dependent on both genetic and environmental factors (Ezaz et al., 2006; Baroiller et al., 2009). Besides hermaphrodite species, two main sex determination system exist: XY and ZW, and they are both present in fish species. It has also been seen that phylogenetically

TABLE 3 | Marker numbers, length, density, and correlations for male, female, and sex-averaged Build 2 linkage map.

LG	No. of SNPs	Physical length[1]	Length (cM)			(F:M length)	Marker density per Mb	Marker density per cM			Correlation[2]		
			F	M	SA			F	M	SA	F	M	SA
LG01	2112	38.37	71.64	53.35	62.11	1.34	55.04	29.48	39.59	34	0.98	0.98	0.99
LG02	1749	35.26	68.22	80.97	66.27	0.84	49.60	25.64	21.6	26.39	0.98	0.99	0.99
LG03	1349	68.55	99.59	68.38	84.91	1.46	19.68	13.55	19.73	15.89	0.91	0.84	0.88
LG04	1707	38.04	60.9	62.85	61.39	0.97	44.87	28.03	27.16	27.81	0.99	0.98	0.99
LG05	1925	34.63	70.2	53.59	61.02	1.31	55.59	27.42	35.92	31.55	0.99	0.99	0.99
LG06	1948	44.57	71.66	80.09	73.6	0.89	43.71	27.18	24.32	26.47	0.99	0.98	0.99
LG07	3391	62.06	132.69	67.45	96.68	1.97	54.64	25.56	50.27	35.07	0.99	0.99	0.99
LG08	1607	30.80	84.18	72.48	77.72	1.16	52.18	19.09	22.17	20.68	0.98	0.98	0.99
LG09	1564	27.52	65.26	58.34	60.39	1.12	56.83	23.97	26.81	25.9	0.97	0.97	0.98
LG10	1387	32.43	63.42	56.51	59.69	1.12	42.77	21.87	24.54	23.24	0.98	0.97	0.98
LG11	1821	36.47	78.49	62.87	70.07	1.25	49.93	23.2	28.96	25.99	0.97	0.99	0.99
LG12	1979	41.23	69.03	56.09	61.99	1.23	48.00	28.67	35.28	31.92	0.98	0.99	0.99
LG13	1614	32.34	72.9	54.64	62.79	1.33	49.91	22.14	29.54	25.7	0.99	0.99	0.99
LG14	2030	39.26	69.67	55.02	61.99	1.27	51.71	29.14	36.9	32.75	0.99	0.98	0.99
LG15	1836	36.15	65	54.95	58.68	1.18	50.79	28.25	33.41	31.29	0.95	0.97	0.96
LG16	1862	43.86	71.61	59.94	64.36	1.19	42.45	26	31.06	28.93	0.99	0.98	0.99
LG17	2005	40.92	68.88	60.36	63.97	1.14	49.00	29.11	33.22	31.34	0.98	0.97	0.98
LG18	1628	37.01	63.14	61.85	62.1	1.02	43.99	25.78	26.32	26.22	0.99	0.99	1
LG19	1646	31.25	64.21	50.37	56.04	1.27	52.67	25.63	32.68	29.37	0.98	0.99	0.98
LG20	1899	36.77	81.26	62.64	71.31	1.3	51.65	23.37	30.32	26.63	0.99	0.99	0.99
LG22	1643	37.01	67.09	72.69	69.25	0.92	44.39	24.49	22.6	23.73	0.98	0.98	0.98
LG23	1484	44.10	73.86	54.17	63.36	1.36	33.65	20.09	27.4	23.42	0.96	0.98	0.97
Total	40186	868.59	1632.9	1359.6	1469.69								
Average	1827	39.48	74.22	61.80	66.80	1.20	47.41	24.61	29.56	27.34	0.98	0.98	0.98

[1]*The physical length (Mb) is retrieved from O_niloticus_UMD1 Assembly.*
[2]*The correlation between the genetic distance of SNPs (cM) on the linkage map and the physical distance (bp) according to the reference genome assembly.*
F, M, and SA represents female, male, and sex-averaged maps, respectively.

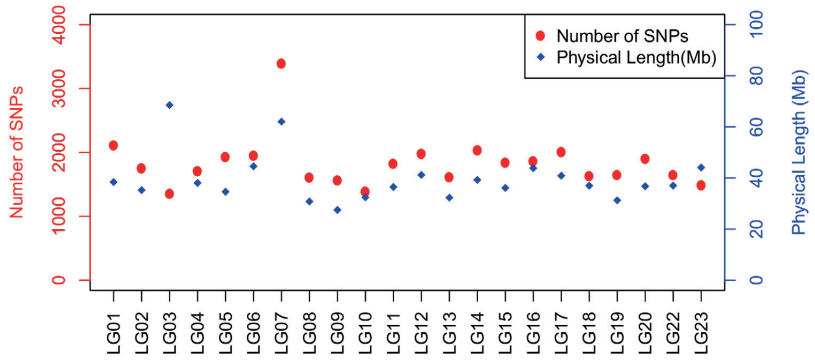


FIGURE 3 | Plot illustrating the number of SNPs and physical length of LG based on O_niloticus_UMD1 Assembly and Build 2 linkage map.

closely related fish species, even in same genus, have different sex determination systems. For example, Blue tilapia, *Oreochromis aureus*, has the ZW sex determination system (Campos-Ramos et al., 2001), where males are homogametic (ZZ) and females are heterogametic (ZW), so the ovum determines the sex of the offspring. On the other hand, Nile tilapia (*O. niloticus*) and Mozambique tilapia (*O. mossambicus*) have the XY system of sex determination, where the males are heterogametic (XY) and females are homogametic (XX), so the sperm determines the sex of the offspring (Mair et al., 1991; Campos-Ramos et al., 2003).

In our study, the sex locus for Nile tilapia was coded using the XY system and mapped to LG23 (**Table 4**) as reported previously in several studies (Karayücel et al., 2004; Shirak et al., 2006; Eshel et al., 2011, 2012). SNP AX-164998274 (SNP probe: AGGTGTGTGGTCTTTCTTTGGAAGTCTGCAGAGT G[C/T]TTCAATAACACAGGTATGGTTTCTCGTTGTGATTC) mapped to the same genetic position as the sex locus. The most

TABLE 4 | Mapping of sex-determination locus in the vicinity of the anti-Müllerian hormone (*amh*) gene.

SNPs/gene	LG	Position (bp)	Male (cM)	Female (cM)	Average (cM)
AX-165032341	LG23	34305951	35.03	44.83	39.97
AX-164990538	LG23	34306186	35.03	44.83	39.97
AX-165017655	LG23	34319855	35.03	44.83	39.97
AX-165032969	LG23	34336514	35.03	44.83	39.97
AX-165012489	LG23	34351488	35.03	44.83	39.97
AX-164995826	LG23	34367182	35.24	44.83	40.00
AX-165001648	LG23	34380102	35.45	44.83	40.09
AX-165030187	LG23	34380282	35.45	44.85	40.12
AX-164992183	LG23	34398468	35.45	44.88	40.13
AX-165006758	LG23	34424845	35.45	44.95	40.15
AX-164986178	LG23	34437472	35.45	45.02	40.20
AX-165024637	LG23	34451454	35.45	45.61	40.53
AX-165013086	LG23	34465412	35.45	45.61	40.53
AX-164998274*	LG23	34496900	35.45	45.61	40.53
amh_delta-y	LG23	34491516-34499598			
amhy	LG23	34491516-34503495			
amh	LG23	34491516-34509687			
AX-164990628	LG23	34510978	35.45	45.61	40.53
AX-165031999	LG23	34511701	35.46	45.61	40.54
AX-165013176	LG23	34525091	35.46	45.61	40.54
AX-165010851	LG23	34576386	35.46	45.61	40.54
AX-164993854	LG23	34585587	35.46	45.61	40.54
AX-164989444	LG23	34598712	35.46	45.61	40.54

*SNP AX-164998274 (marked as *) mapped to the same genetic position as the Phenotypic sex of the individuals in the Build 1 linkage map.*

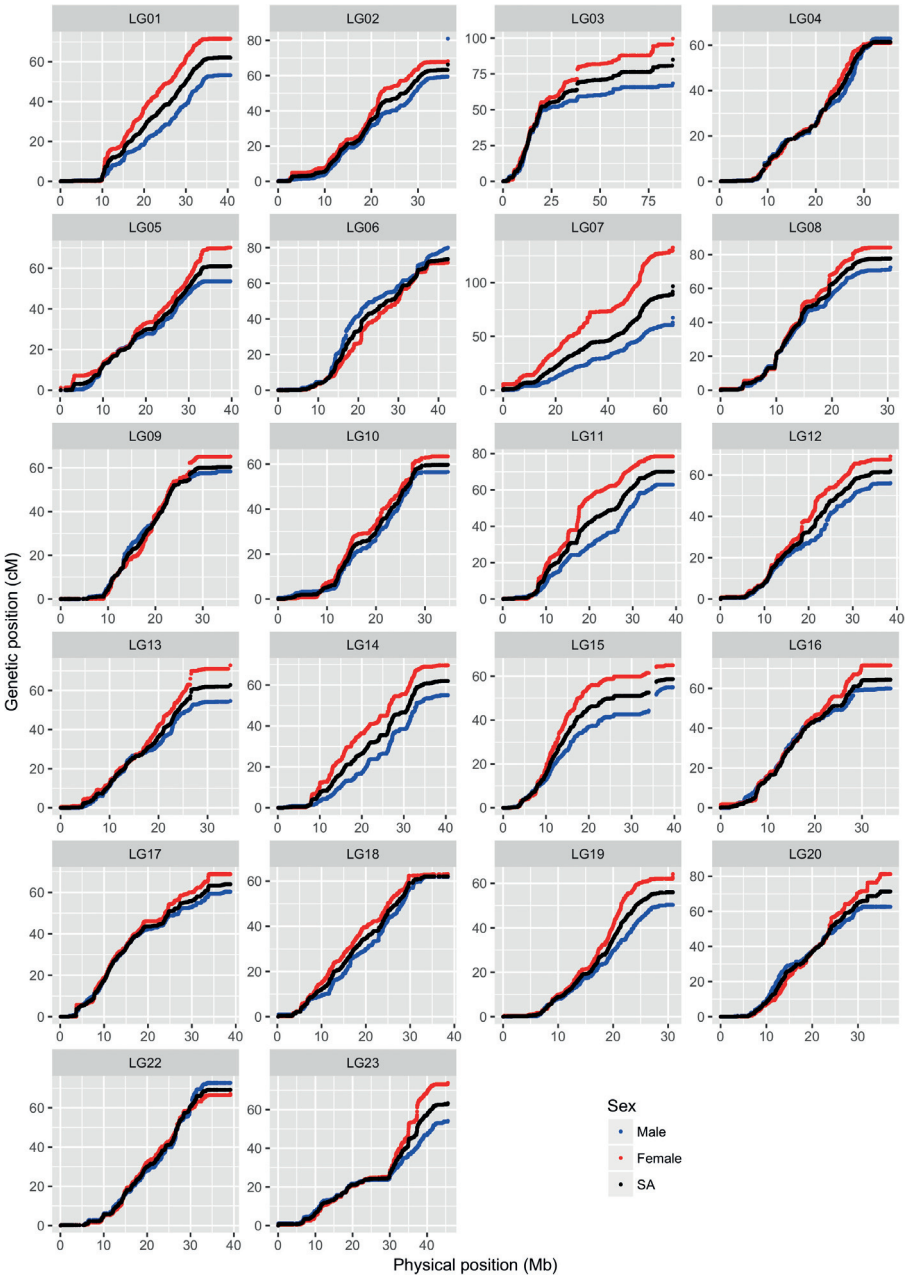


FIGURE 4 | Comparison of map positions between genetic and physical maps for different LGs in Build 2. The *y*-axis gives the linkage map positions, and the *x*-axis gives the physical positions. Linkage groups (LGs) and the physical positions are based on O_niloticus_UMD_NMBU Assembly. The maps are color-coded: red for female specific, blue for male specific, and black for sex-averaged linkage maps. Please note that the 22 LGs in the Nile tilapia have been named from LG01 to LG23 (no LG21).

likely position of sex locus (pos. 34.5 Mb/40.53 cM on LG23) maps close to the anti-Müllerian hormone (*amh*) gene, previously characterized as sex determining gene in Nile tilapia (Li et al., 2015).

Implications in Tilapia Industry

Tilapia is a commercially important aquaculture species, with more than 3.9 million tons of fish and filets being traded in 2015 (FAO, 2017) and more than 20 breeding programs (Neira, 2010). The present SNP array and linkage map has the potential to greatly improve the genetic gain for this economic important species, and help surpass the difficulties of efficient selection for the invasively measured traits, the traits which cannot be measured directly on the candidate broodstock fish, but are only measured on the sibs of the candidates, e.g., disease resistance, filet yield, etc. These tools may also be useful to bridge the genotype-phenotype gap in Nile tilapia, which has been pursued for a long time (Gjøen, 2004).

A major capability of these resources will be to find economic important QTLs or chromosome regions affecting economically important traits like disease resistance, filet traits or feed efficiency. In order to fine map these QTLs, it is essential to have a high-resolution linkage map. The dense linkage map can also be integrated with physical maps to position and orient scaffolds along LGs, thereby producing genome assemblies of higher quality.

Another important implication will be to facilitate the shift from traditional breeding strategies to genomic selection in Nile tilapia. In the future, breeding goals in Nile tilapia will include many invasively measured traits. Genomic selection will significantly help us to overcome these challenges, increasing the profitability and the genetic gain (Meuwissen et al., 2001; Nielsen et al., 2009; Sonesson and Meuwissen, 2009; Vela-Avitúa et al., 2015; Hosoya et al., 2017; Houston, 2017). Finally, this will also help to discern more accurately the additive from the non-additive genetic effects, thereby increasing the selection accuracy and the possibility to utilize non-additive genetic effects (Varona et al., 2018).

Another obvious use of the SNP-array will be in the parentage assignments. The drawback of the conventional breeding designs in Nile tilapia using PIT tags is the confounding of the full-sib family effects (due to communal rearing of full-sibs) and maternal environmental effects (due to mouth brooding), making it difficult to detangle the various variance components accurately (Joshi et al., 2018), which ultimately decreases the accuracy of the selection.

CONCLUSION

We present the first SNP-array, the Onil50-array, containing ca 58,000 SNPs for Nile tilapia, which was validated in close to 5000 individuals. Further, we constructed a high density integrated genetic and physical linkage map, with LGs showing sex-differentiated sigmoidal recombination patterns. These new resources has the potential to greatly influence and improve the genetic gain when applying genomic selection and surpass the difficulties of efficient selection for invasively measured traits in Nile tilapia.

DATA AVAILABILITY

The assemblies used in this study can be found in NCBI using the following accessions: Orenil1.1 = GCF_000188235.2, O_niloticus_UMD1 = MKQE00000000, and O_niloticus_UMD_NMBU = MKQE02000000. The whole genome sequence data used for SNP detection has been deposited in the European Nucleotide Archive and can be found in EMBL-EBI_website using accession PRJEB28330. Linkage map generated from this study can be found in the Figshare: https://figshare.com/s/8427b97cf6e623173232.

ETHICS STATEMENT

Sampling of DNA was done in accordance with the commercial practice and norms by Genomar Genetics AS.

AUTHOR CONTRIBUTIONS

HG, AA, and MK conceived and designed the study. AA coordinated biological sampling. MK and MÁ were responsible for array design and MÁ performed lab work and initial analysis of results. RJ constructed the linkage map, while SL integrated the genetic and physical maps. RJ and MÁ prepared the draft manuscript which was reviewed and edited by HG, MK, AA, and SL. All authors read and approved the manuscript.

ACKNOWLEDGMENTS

We would like to acknowledge Anders Skaarud from Genomar Genetics AS for his help in sample management. We would also like to thank Harald Grove, Torfinn Nome, and Tim Knutsen for their valuable advice and help with bioinformatics analyses of sequence and SNP data. The SNP array was developed in cooperation with Affymetrix, Inc., and we particularly thank the following Affymetrix personnel for their direct contribution: Lakshmi Radhakrishnan and Alessandro Davassi. We also thank Silje Karoliussen for her help in genotyping and raw data analyses. Similarly, we also acknowledge Solomon Boison and Luqman Aslam for their interaction during linkage map construction.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: https://www.frontiersin.org/articles/10.3389/fgene.2018.00472/full#supplementary-material

REFERENCES

ADB (2005). *An Impact Evaluation of the Development of Genetically Improved Farmed Tilapia and their Dissemination in Selected Countries*. Mandaluyong: Asian Development Bank.

Adenyo, C., Ogden, R., Kayang, B., Onuma, M., Nakajima, N., and Inoue-Murayama, M. (2017). Genome-wide DNA markers to support genetic management for domestication and commercial production in a large rodent, the Ghanaian grasscutter (*Thryonomys swinderianus*). *Anim. Genet.* 48, 113–115. doi: 10.1111/age.12478

Ayllon, F., Kjærner-Semb, E., Furmanek, T., Wennevik, V., Solberg, M. F., Dahle, G., et al. (2015). The vgll3 locus controls age at maturity in wild and domesticated Atlantic salmon (*Salmo salar* L.) males. *PLoS Genet.* 11:e1005628. doi: 10.1371/journal.pgen.1005628

Bangera, R., Correa, K., Lhorente, J. P., Figueroa, R., and Yáñez, J. M. (2017). Genomic predictions can accelerate selection for resistance against *Piscirickettsia salmonis* in Atlantic salmon (*Salmo salar*). *BMC Genomics* 18:121. doi: 10.1186/s12864-017-3487-y

Baroiller, J.-F., D'Cotta, H., Bezault, E., Wessels, S., and Hoerstgen-Schwark, G. (2009). Tilapia sex determination: where temperature and genetics meet. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 153, 30–38. doi: 10.1016/j.cbpa.2008.11.018

Barría, A., Christensen, K. A., Yoshida, G. M., Correa, K., Jedlicki, A., Lhorente, J. P., et al. (2018). Genomic predictions and genome-wide association study of resistance against *Piscirickettsia salmonis* in coho salmon (*Oncorhynchus kisutch*) using ddRAD sequencing. *G3* 8, 1183–1194. doi: 10.1534/g3.118.200053

Campos-Ramos, R., Harvey, S. C., Masabanda, J. S., Carrasco, L. A., Griffin, D. K., McAndrew, B. J., et al. (2001). Identification of putative sex chromosomes in the blue tilapia, *Oreochromis aureus*, through synaptonemal complex and FISH analysis. *Genetica* 111, 143–153. doi: 10.1023/A:1013707818534

Campos-Ramos, R., Harvey, S. C., McAndrew, B. J., and Penman, D. J. (2003). An investigation of sex determination in the Mozambique tilapia, *Oreochromis mossambicus*, using synaptonemal complex analysis, FISH, sex reversal and gynogenesis. *Aquaculture* 221, 125–140. doi: 10.1016/S0044-8486(03)00072-3

Chang, C. C., Chow, C. C., Tellier, L. C. A. M., Vattikuti, S., Purcell, S. M., and Lee, J. J. (2015). Second-generation PLINK: rising to the challenge of larger and richer datasets. *Gigascience* 4:7. doi: 10.1186/s13742-015-0047-8

Cingolani, P., Platts, A., Wang, L. L., Coon, M., Nguyen, T., Wang, L., et al. (2012). A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of *Drosophila melanogaster* strain w1118; iso-2; iso-3. *Fly* 6, 80–92. doi: 10.4161/fly.19695

Conte, M. A., Gammerdinger, W. J., Bartie, K. L., Penman, D. J., and 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. doi: 10.1186/s12864-017-3723-5

Conte, M. A., Joshi, R., Moore, E. C., Nandamuri, S. P., Gammerdinger, W. J., Clark, F. E., et al. (2018). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *bioRxiv* [Preprint]. doi: 10.1101/383992

Coop, G., and Przeworski, M. (2007). An evolutionary view of human recombination. *Nat. Rev. Genet.* 8, 23–34. doi: 10.1038/nrg1947

Correa, K., Bangera, R., Figueroa, R., Lhorente, J. P., and Yáñez, J. M. (2017a). The use of genomic information increases the accuracy of breeding value predictions for sea louse (*Caligus rogercresseyi*) resistance in Atlantic salmon (*Salmo salar*). *Genet. Sel. Evol.* 49:15. doi: 10.1186/s12711-017-0291-8

Correa, K., Lhorente, J. P., Bassini, L., López, M. E., Di Genova, A., Maass, A., et al. (2017b). Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar* L.) using a 50K SNP genotyping array. *Aquaculture* 472, 61–65. doi: 10.1016/j.aquaculture.2016.04.008

Correa, K., Lhorente, J. P., López, M. E., Bassini, L., Naswa, S., Deeb, N., et al. (2015). Genome-wide association analysis reveals loci associated with resistance against *Piscirickettsia salmonis* in two Atlantic salmon (*Salmo salar* L.) chromosomes. *BMC Genomics* 16:854. doi: 10.1186/s12864-015-2038-7

Davidson, W. S., and Yáñez, J. M. (2016). Genome wide association study for resistance to *Caligus rogercresseyi* in Atlantic salmon (*Salmo salar* L.) using a 50K SNP genotyping array. *Aquaculture* 472, 61–65. doi: 10.1016/j.aquaculture.2016.04.008

Eknath, A. E., Tayamen, M. M., Palada-de Vera, M. S., Danting, J. C., Reyes, R. A., Dionisio, E. E., et al. (1993). Genetic improvement of farmed tilapias: the growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments. *Aquaculture* 111, 171–188. doi: 10.1016/B978-0-444-81527-9.50021-X

Eshel, O., Shirak, A., Weller, J. I., Hulata, G., and Ron, M. (2012). Linkage and physical mapping of sex region on LG23 of Nile tilapia (*Oreochromis niloticus*). *G3* 2, 35–42. doi: 10.1534/g3.111.001545

Eshel, O., Shirak, A., Weller, J. I., Slossman, T., Hulata, G., Cnaani, A., et al. (2011). Fine-mapping of a locus on linkage group 23 for sex determination in Nile tilapia (*Oreochromis niloticus*). *Anim. Genet.* 42, 222–224. doi: 10.1111/j.1365-2052.2010.02128.x

Ezaz, T., Stiglec, R., Veyrunes, F., and Graves, J. A. M. (2006). Relationships between vertebrate ZW and XY sex chromosome systems. *Curr. Biol.* 16, R736–R743. doi: 10.1016/j.cub.2006.08.021

FAO (2017). *FAO Yearbook. Fishery and Aquaculture Statistics. 2015*. Rome: FAO.

Ferreira, I. A., Poletto, A. B., Kocher, T. D., Mota-Velasco, J. C., Penman, D. J., and Martins, C. (2010). Chromosome evolution in African cichlid fish: contributions from the physical mapping of repeated DNAs. *Cytogenet. Genome Res.* 129, 314–322. doi: 10.1159/000315895

Fujimura, K., and Okada, N. (2007). Development of the embryo, larva and early juvenile of Nile tilapia *Oreochromis niloticus* (Pisces: Cichlidae). Developmental staging system. *Dev. Growth Differ.* 49, 301–324. doi: 10.1111/j.1440-169X.2007.00926.x

Garrison, E., and Marth, G. (2012). Haplotype-based variant detection from short-read sequencing. *arXiv* [Preprint]. https://arxiv.org/abs/1207.3907

Gjøen, H. M. (2004). "A new era: the merging of quantitative and molecular genetics—Prospects for tilapia breeding programs," in *Proceedings of the 6th International Symposium on Tilapia in Aquaculture*, eds R. B. Bolivar, G. C. Mair, and K. Fitzsimmons (Manila: Bureau of Fisheries and Aquatic Resources).

Gonzalez-Pena, D., Gao, G., Baranski, M., Moen, T., Cleveland, B. M., Kenney, P. B., et al. (2016). Genome-wide association study for identifying loci that affect fillet yield, carcass, and body weight traits in rainbow trout (*Oncorhynchus mykiss*). *Front. Genet.* 7:203. doi: 10.3389/fgene.2016.00203

Graffelman, J., Jain, D., and Weir, B. (2017). A genome-wide study of Hardy–Weinberg equilibrium with next generation sequence data. *Hum. Genet.* 136, 727–741. doi: 10.1007/s00439-017-1786-7

Gruhn, J. R., Rubio, C., Broman, K. W., Hunt, P. A., and Hassold, T. (2013). Cytological studies of human meiosis: sex-specific differences in recombination originate at, or prior to, establishment of double-strand breaks. *PLoS One* 8:e85075. doi: 10.1371/journal.pone.0085075

Gutierrez, A. P., Yáñez, J. M., and Davidson, W. S. (2016). Evidence of recent signatures of selection during domestication in an Atlantic salmon population. *Mar. Genomics* 26, 41–50. doi: 10.1016/j.margen.2015.12.007

Gutierrez, A. P., Yáñez, J. M., Fukui, S., Swift, B., and 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. doi: 10.1371/journal.pone.0119730

Guyon, R., Rakotomanga, M., Azzouzi, N., Coutanceau, J. P., Bonillo, C., D'Cotta, H., et al. (2012). A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* 13:222. doi: 10.1186/1471-2164-13-222

Hayes, B. J. (2011). Technical note: efficient parentage assignment and pedigree reconstruction with dense single nucleotide polymorphism data. *J. Dairy Sci.* 94, 2114–2117. doi: 10.3168/jds.2010-3896

Hosoya, S., Kikuchi, K., Nagashima, H., Onodera, J., Sugimoto, K., Satoh, K., et al. (2017). Genomic selection in aquaculture. *Bull. Jap. Fish. Res. Educ. Agen.* 45, 35–39.

Houston, R. D. (2017). Future directions in breeding for disease resistance in aquaculture species. *Rev. Bras. Zootec.* 46, 545–551. doi: 10.1590/s1806-92902017000600010

Joshi, R., Woolliams, J., Meuwissen, T., and Gjøen, H. (2018). Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity* 120, 452–462. doi: 10.1038/s41437-017-0046-x

Karayücel, I., Ezaz, T., Karayücel, S., McAndrew, B. J., and Penman, D. J. (2004). Evidence for two unlinked "sex reversal" loci in the Nile tilapia, *Oreochromis niloticus*, and for linkage of one of these to the red body colour gene. *Aquaculture* 234, 51–63. doi: 10.1016/j.aquaculture.2003.12.016

Kocher, T. D., Lee, W. J., Sobolewska, H., Penman, D., and McAndrew, B. (1998). A genetic linkage map of a cichlid fish, the tilapia (*Oreochromis niloticus*). *Genetics* 148, 1225–1232.

Lee, B.-Y., Hulata, G., and Kocher, T. D. (2004). Two unlinked loci controlling the sex of blue tilapia (*Oreochromis aureus*). *Heredity* 92, 543–549. doi: 10.1038/sj.hdy.6800453

Lee, B.-Y., Lee, W.-J., Streelman, J. T., Carleton, K. L., Howe, A. E., Hulata, G., et al. (2005). A second-generation genetic linkage map of tilapia (*Oreochromis* spp.). *Genetics* 170, 237–244. doi: 10.1534/genetics.104.035022

Lee, S., Kasif, S., Weng, Z., and Cantor, C. R. (2008). Quantitative analysis of single nucleotide polymorphisms within copy number variation. *PLoS One* 3:e3906. doi: 10.1371/journal.pone.0003906

Lenormand, T., and Dutheil, J. (2005). Recombination difference between sexes: a role for haploid selection. *PLoS Biol.* 3:e63. doi: 10.1371/journal.pbio.0030063

Li, H. (2013). Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv* [Preprint]. https://arxiv.org/abs/1303.3997

Li, H., Handsaker, B., Wysoker, A., Fennell, T., Ruan, J., Homer, N., et al. (2009). The sequence alignment/Map format and SAMtools. *Bioinformatics* 25, 2078–2079. doi: 10.1093/bioinformatics/btp352

Li, M., Sun, Y., Zhao, J., Shi, H., Zeng, S., Ye, K., et al. (2015). A tandem duplicate of anti-müllerian hormone with a missense SNP on the Y chromosome is essential for male sex determination in Nile tilapia, *Oreochromis niloticus*. *PLoS Genet.* 11:e1005678. doi: 10.1371/journal.pgen.1005678

Li, Y., Liu, S., Qin, Z., Waldbieser, G., Wang, R., Sun, L., et al. (2014). Construction of a high-density, high-resolution genetic map and its integration with BAC-based physical map in channel catfish. *DNA Res.* 22, 39–52. doi: 10.1093/dnares/dsu038

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

Liu, S., Vallejo, R. L., Palti, Y., Gao, G., Marancik, D. P., Hernandez, A. G., et al. (2015). Identification of single nucleotide polymorphism markers associated with bacterial cold water disease resistance and spleen size in rainbow trout. *Front. Genet.* 6:298. doi: 10.3389/fgene.2015.00298

Mair, G. C., Scott, A. G., Penman, D. J., Beardmore, J. A., and Skibinski, D. O. F. (1991). Sex determination in the genus *Oreochromis*: 1. Sex reversal, gynogenesis and triploidy in *O. niloticus*. *Theor. Appl. Genet.* 82, 144–152. doi: 10.1007/BF00226205

Majumdar, K. C., and McAndrew, B. J. (1986). Relative DNA content of somatic nuclei and chromosomal studies in three genera, *Tilapia*, *Sarotherodon*, and *Oreochromis* of the tribe Tilapiini (Pisces, Cichlidae). *Genetica* 68, 175–188. doi: 10.1007/BF02424441

Meuwissen, T. H. E., Hayes, B. J., and Goddard, M. E. (2001). Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157, 1819–1829.

NCBI (2018). *Assembly Information by Organism*. Available at: https://www.ncbi.nlm.nih.gov/assembly/organism/8128/all/

Neira, R. (2010). "Breeding in aquaculture species: genetic improvement programs in developing countries," in *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Leipzig.

Nielsen, H. M., Sonesson, A. K., Yazdi, H., and Meuwissen, T. H. E. (2009). Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. *Aquaculture* 289, 259–264. doi: 10.1016/j.aquaculture.2009.01.027

Ødegård, J., Moen, T., Santi, N., Korsvoll, S. A., Kjøglum, S., and Meuwissen, T. H. E. (2014). Genomic prediction in an admixed population of Atlantic salmon (*Salmo salar*). *Front. Genet.* 5:402. doi: 10.3389/fgene.2014.00402

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

Palaiokostas, C., Kocour, M., Prchal, M., and Houston, R. D. (2018). Accuracy of genomic evaluations of juvenile growth rate in common carp (*Cyprinus carpio*) using genotyping by sequencing. *Front. Genet.* 9:82. doi: 10.3389/fgene.2018.00082

Palti, Y., Vallejo, R. L., Gao, G., Liu, S., Hernandez, A. G., Rexroad, C. E. III, et al. (2015). Detection and validation of QTL affecting bacterial cold water disease resistance in rainbow trout using restriction-site associated DNA sequencing. *PLoS One* 10:e0138435. doi: 10.1371/journal.pone.0138435

Poletto, A. B., Ferreira, I. A., Cabral-de-Mello, D. C., Nakajima, R. T., Mazzuchelli, J., Ribeiro, H. B., et al. (2010). Chromosome differentiation patterns during cichlid fish evolution. *BMC Genet.* 11:50. doi: 10.1186/1471-2156-11-50

Rastas, P., Paulin, L., Hanski, I., Lehtonen, R., and Auvinen, P. (2013). Lep-MAP: fast and accurate linkage map construction for large SNP datasets. *Bioinformatics* 29, 3128–3134. doi: 10.1093/bioinformatics/btt563

Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P., and Houston, R. (2017). Applications of genotyping by sequencing in aquaculture breeding and genetics. *Rev. Aquac.* 10, 670–682. doi: 10.1111/raq.12193

Roesti, M., Moser, D., and Berner, D. (2013). Recombination in the threespine stickleback genome—patterns and consequences. *Mol. Ecol.* 22, 3014–3027. doi: 10.1111/mec.12322

Seale, A. P., Riley, L. G., Leedom, T. A., Kajimura, S., Dores, R. M., Hirano, T., et al. (2002). Effects of environmental osmolality on release of prolactin, growth hormone and ACTH from the tilapia pituitary. *Gen. Comp. Endocrinol.* 128, 91–101. doi: 10.1016/S0016-6480(02)00027-8

Shirak, A., Seroussi, E., Cnaani, A., Howe, A. E., Domokhovsky, R., Zilberman, N., et al. (2006). Amh and Dmrta2 genes map to tilapia (*Oreochromis* spp.) linkage group 23 within quantitative trait locus regions for sex determination. *Genetics* 174, 1573–1581. doi: 10.1534/genetics.106.059030

Sodeland, M., Gaarder, M., Moen, T., Thomassen, M., Kjøglum, S., Kent, M., et al. (2013). Genome-wide association testing reveals quantitative trait loci for fillet texture and fat content in Atlantic salmon. *Aquaculture* 408, 169–174. doi: 10.1016/j.aquaculture.2013.05.029

Sonesson, A. K., and Meuwissen, T. H. E. (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet. Sel. Evol.* 41:37. doi: 10.1186/1297-9686-41-37

Thermo Fisher Scientific Inc (2018). *Axiom™Analysis Suite (AxAS) v4.0 USER GUIDE*. Available at: https://downloads.thermofisher.com/Affymetrix_Softwares/Axiom_Analysis_Suite_AxAS_v4.0_User_Guide.pdf

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Bron, J. E., Taggart, J. B., et al. (2016). Genomic prediction of host resistance to sea lice in farmed Atlantic salmon populations. *Genet. Sel. Evol.* 48:47. doi: 10.1186/s12711-016-0226-9

Tsai, H. Y., Hamilton, A., Guy, D. R., Tinch, A. E., Bishop, S. C., and Houston, R. D. (2015a). Verification of SNPs associated with growth traits in two populations of farmed Atlantic salmon. *Int. J. Mol. Sci.* 17:5. doi: 10.3390/ijms17010005

Tsai, H.-Y., Hamilton, A., Tinch, A. E., Guy, D. R., Gharbi, K., Stear, M. J., et al. (2015b). Genome wide association and genomic prediction for growth traits in juvenile farmed Atlantic salmon using a high density SNP array. *BMC Genomics* 16:969. doi: 10.1186/s12864-015-2117-9

Vallejo, R., Leeds, T., Liu, S., Gao, G., Welch, T., Wiens, G., et al. (2015a). "Accuracy of genomic prediction for BCWD resistance in rainbow trout using different genotyping platforms and genomic selection models," in *Proceedings of the Plant and Animal Genome Conference*, San Diego, CA, 726.

Vallejo, R., Leeds, T., Liu, S., Gao, G., Welch, T., Wiens, G., et al. (2015b). "Genomic selection for BCWD resistance in Rainbow trout using RADSNP and SNP genotyping platforms, single-step GBLUP and Bayesian variable selection models," in *Proceedings of the International Symposium on Genetics in Aquaculture*, Townsville, 05796.

Vallejo, R. L., Liu, S., Gao, G., Fragomeni, B. O., Hernandez, A. G., Leeds, T. D., et al. (2017a). Similar genetic architecture with shared and unique quantitative trait loci for bacterial cold water disease resistance in two rainbow trout breeding populations. *Front. Genet.* 8:156. doi: 10.3389/fgene.2017.00156

Vallejo, R. L., Leeds, T. D., Gao, G., Parsons, J. E., Martin, K. E., Evenhuis, J. P., et al. (2017b). Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. *Genet. Sel. Evol.* 49:17. doi: 10.1186/s12711-017-0293-6

Varona, L., Legarra, A., Toro, M. A., and Vitezica, Z. G. (2018). Non-additive effects in genomic selection. *Front. Genet.* 9:78. doi: 10.3389/fgene.2018.00078

Vela-Avitúa, S., Meuwissen, T. H. E., Luan, T., and Ødegård, J. (2015). Accuracy of genomic selection for a sib-evaluated trait using identity-by-state and identity-by-descent relationships. *Genet. Sel. Evol.* 47:9. doi: 10.1186/s12711-014-0084-2

Vilela, D. A. R., Silva, S. G. B., Peixoto, M. T. D., Godinho, H. P., and França, L. R. (2003). Spermatogenesis in teleost: insights from the Nile tilapia (*Oreochromis niloticus*) model. *Fish Physiol. Biochem.* 28, 187–190. doi: 10.1023/B:FISH.0000030523.16010.62

Wang, L., Wan, Z. Y., Bai, B., Huang, S. Q., Chua, E., Lee, M., et al. (2015). Construction of a high-density linkage map and fine mapping of QTL for growth in Asian seabass. *Sci. Rep.* 5:16358. doi: 10.1038/srep16358

Wright, P. A., and Land, M. D. (1998). Urea production and transport in teleost fishes. *Comp. Biochem. Physiol. Part A Mol. Integr. Physiol.* 119, 47–54. doi: 10.1016/S1095-6433(97)00407-8

Yoshida, G. M., Bangera, R., Carvalheiro, R., Correa, K., Figueroa, R., Lhorente, J. P., et al. (2017a). Genomic prediction accuracy for resistance against *Piscirickettsia salmonis* in farmed rainbow trout. *G3* 8, 719–726. doi: 10.1534/g3.117.300499

Yoshida, G. M., Lhorente, J. P., Carvalheiro, R., and Yáñez, J. M. (2017b). Bayesian genome-wide association analysis for body weight in farmed Atlantic salmon (*Salmo salar* L.). *Anim. Genet.* 48, 698–703. doi: 10.1111/age.12621

Yoshida, G. M., Carvalheiro, R., Rodríguez, F. H., Lhorente, J. P., and Yáñez, J. M. (2018). Single-step genomic evaluation improves accuracy of breeding value predictions for resistance to infectious pancreatic necrosis virus in rainbow trout. *Genomics* doi: 10.1016/j.ygeno.2018.01.008 [Epub ahead of print].

Conflict of Interest Statement: Genomar Genetics AS employs one of the authors, AA.

The remaining authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Supplementary Material

Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (O. niloticus)

Rajesh Joshi[1,*], Mariann Árnyasi[1], Sigbjørn Lien[1], Hans Magnus Gjøen[1], Alejandro Tola Alvarez[2], Matthew Peter Kent[1]

[1]Department of Animal and Aquacultural Sciences, Faculty of Biosciences, Norwegian University of Life Sciences, N-1432, Ås, Norway

[2]Genomar Genetics AS, Norway

*** Correspondence:** Rajesh Joshi: rajesh.joshi@nmbu.no

Supplementary Table 1: Observations in each factorial mating in Group 1 population. 11 different sires (M1 to M11) are mated with 8 different dams (F1 to F8) in factorial manner. Only full-sib families ≥ 8 offspring are shown.

	F1	**F2**	**F3**	**F4**	**F5**	**F6**	**F7**	**F8**	**Total**
M1	22	13	10	-	8	-	-	-	**53**
M2	-	-	-	10	-	21	10	24	**65**
M3	-	-	-	12	-	23	11	16	**62**
M4	-	-	-	13	-	9	-	-	**22**
M5	11	24	9	-	8	-	-	-	**52**
M6	-	-	-	-	-	18	8	13	**39**
M7	19	12	-	-	-	-	-	-	**31**
M8	-	-	-	14	-	9	12	28	**63**
M9	-	14	-	-	10	-	-	-	**24**
M10	16	14	8	-	8	-	-	-	**46**
M11	-	-	-	11	-	-	-	-	**11**
Total	**68**	**77**	**27**	**60**	**34**	**80**	**41**	**81**	**468**

Supplementary Table 2: Observations in different full-sib families in the Group 2 population.

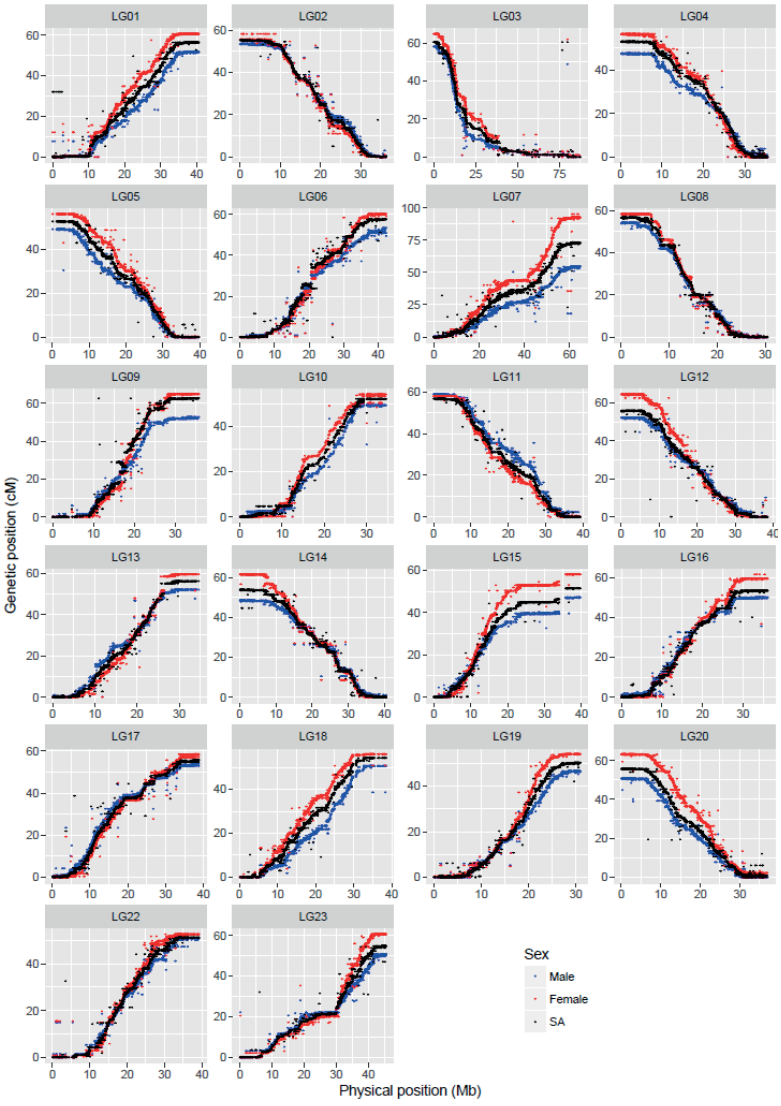
Dam	F9	F10	F11	F12	F13	F14	F15	**Total**
Sire	M12	M13	M14	M15	M16	M17	M18	
No. of offspring	22	24	26	36	37	37	39	**221**

Supplementary Table 3: Published linkage maps for Tilapia species

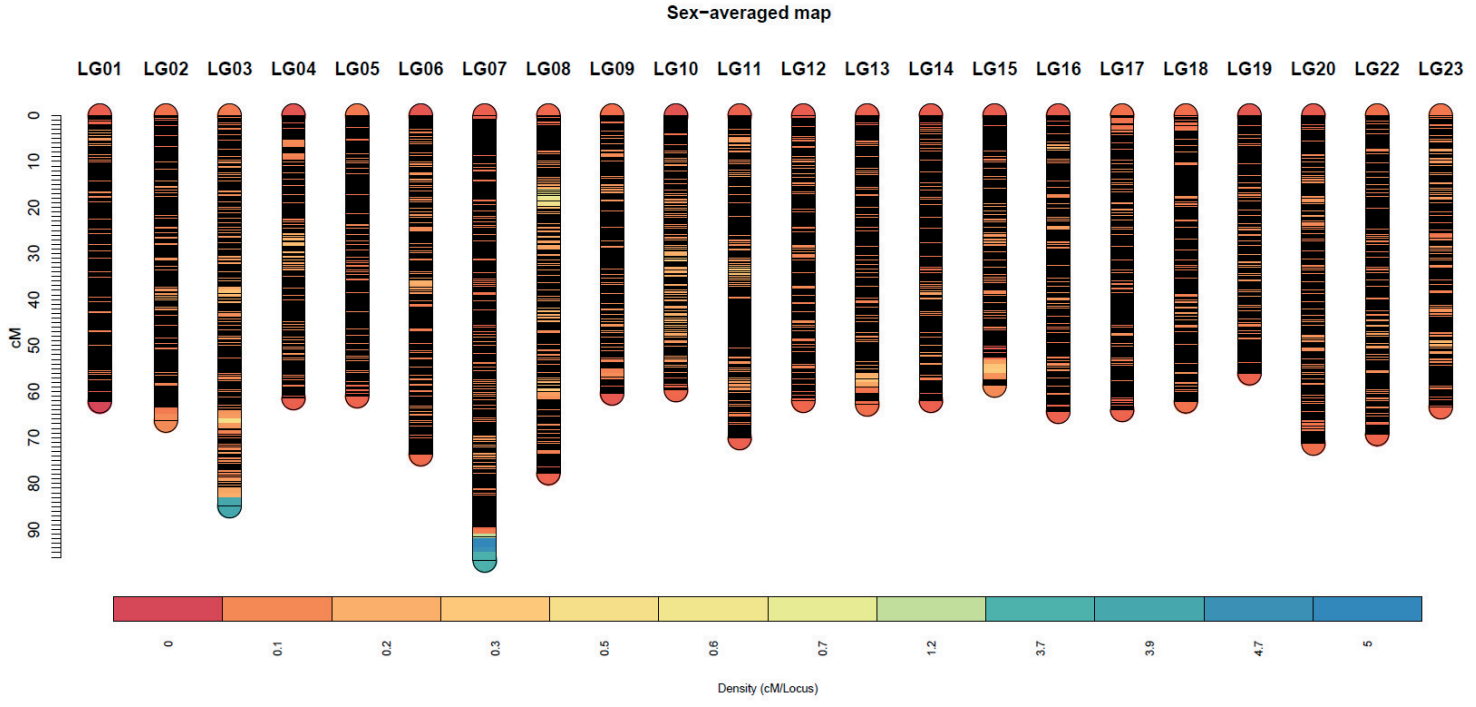
Species of Tilapia		Map length (cM)	Marker number and type	Average marker interval (cM)	Authors & Year
Oreochromis niloticus		704	62 microsatellites + 112 AFLP	-	(Kocher et al., 1998)
O. niloticus X *O. aureus*		1,311	525 microsatellite and 21 gene-based markers	2.4	(Lee et al., 2005)
O. niloticus		34,084 cR_{3500} and 937,310 kb	1358 markers – radiation hybrid (RH) map	742 Kb	(Guyon et al., 2012)
O. niloticus		1,176	3,802 SNPs	0.7	(Palaiokostas et al., 2013)
O. mossambicus	Female	514	13 microsatellites and 49 AFLPs	8.3	(Agresti *et al.*, 2000)
	Male	1632	60 microsatellites and 154 AFLPs	7.6	
O. mossambicus		1042.5	301 markers		(Liu *et al.*, 2013)
O. mossambicus X *O. spp.* (Saline tilapia)	Consensus	1067.6	401 microsatellites including 282 EST-derived markers	3.3	
	Male	950.8	261 markers	3.6	
	Female	1030.6	261 markers	4	
Red tilapia		984.0	320 markers	3.1	

Supplementary Table 4: Summary statistics showing the inter-marker distance (base pairs) of the SNPs on the Onil50-array.

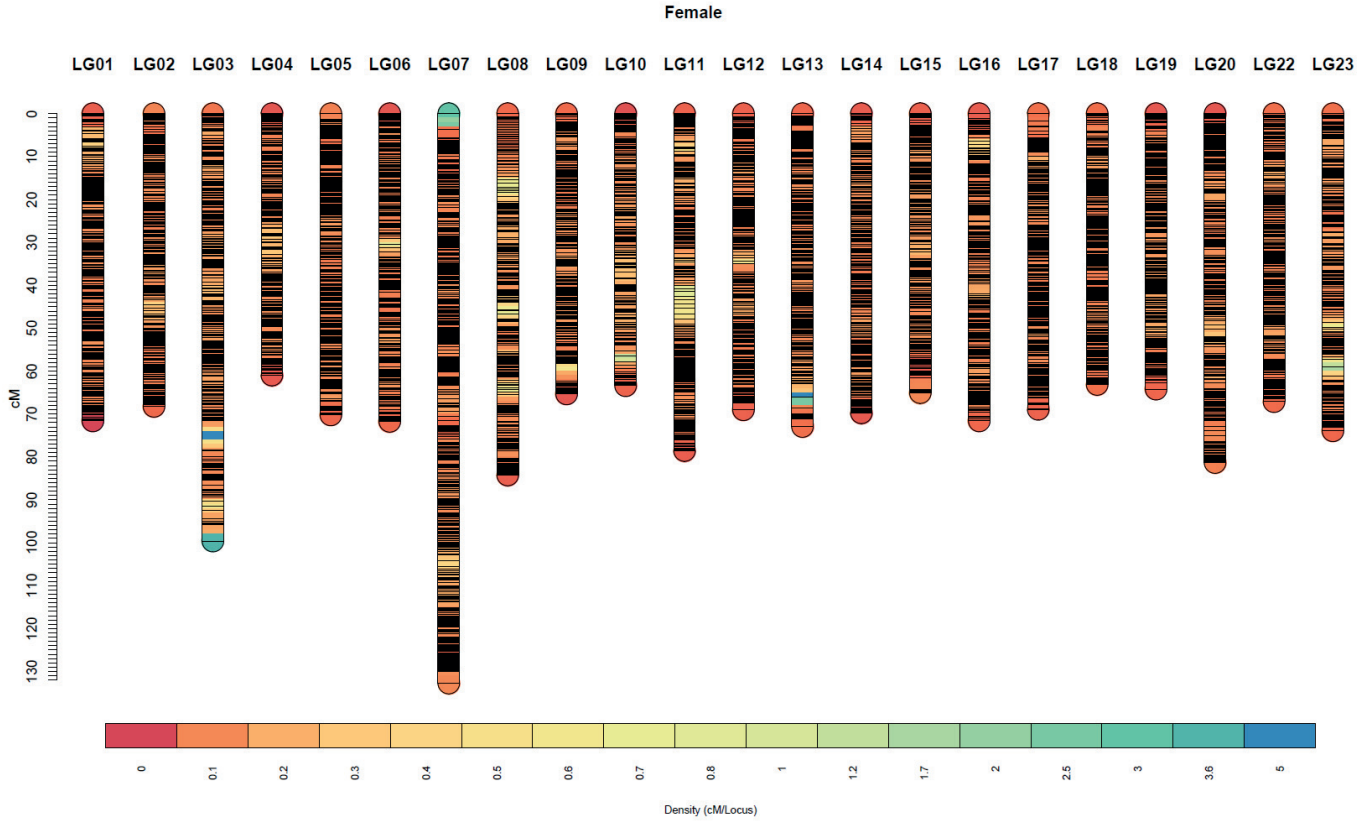
Orenil1.1 assembly			O_niloticus_UMD1 assembly		
LG	**Mean**	**sd**	**LG**	**Mean**	**sd**
LG01	12124	4821	**LG01**	13548	10950
LG02	12253	6180	**LG02**	14666	14189
LG03	13646	10919	**LG03a**	20689	35557
			LG03b	38045	55867
LG04	12522	7061	**LG04**	15509	16232
LG05	12771	6860	**LG05**	13564	8909
LG06	12701	6249	**LG06**	15149	15972
LG07	12356	5419	**LG07**	13227	9624
LG08-24	12702	7093	**LG08**	13326	10639
LG09	12091	5667	**LG09**	14377	16435
LG10	12068	4742	**LG10**	17227	16536
LG11	12608	6704	**LG11**	13642	11567
LG12	12595	6466	**LG12**	14516	14585
LG13	12377	5895	**LG13**	14188	14768
LG14	12652	6626	**LG14**	14634	13260
LG15	12239	6557	**LG15**	15991	13433
LG16-21	12556	6226	**LG16**	15323	17131
LG17	12154	5881	**LG17**	14153	12842
LG18	12623	7246	**LG18**	15992	22063
LG19	12196	5849	**LG19**	13514	12128
LG20	12632	6502	**LG20**	13910	11003
LG22	12603	7125	**LG22**	16212	18656
LG23	12934	8374	**LG23**	19813	22055
Across genome	12505	6538	**Across genome**	15410	17712
Scaffolds	36493	22508	**Scaffolds**	28097	31755
mito	1636	1057	**mito**	1636	1057



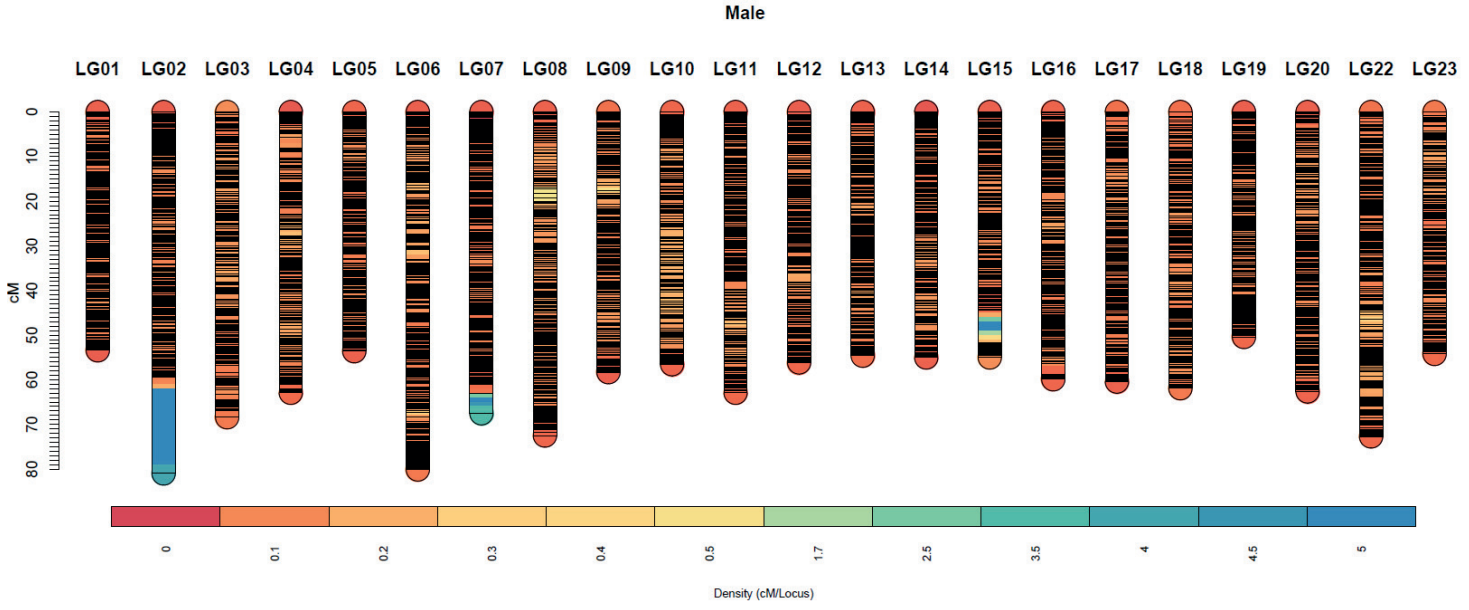
Supplementary Figure 1: Comparison of map positions between genetic and physical maps for different LGs in Build1 linkage map. The y-axis gives the linkage map positions, and the x-axis gives the physical positions. Linkage groups and the physical positions are based on O_niloticus_UMD1 Assembly. The maps are color-coded: red for female specific, blue for male specific and black for sex-averaged linkage maps. Inversion in maps shows that the genetic order is inverted.



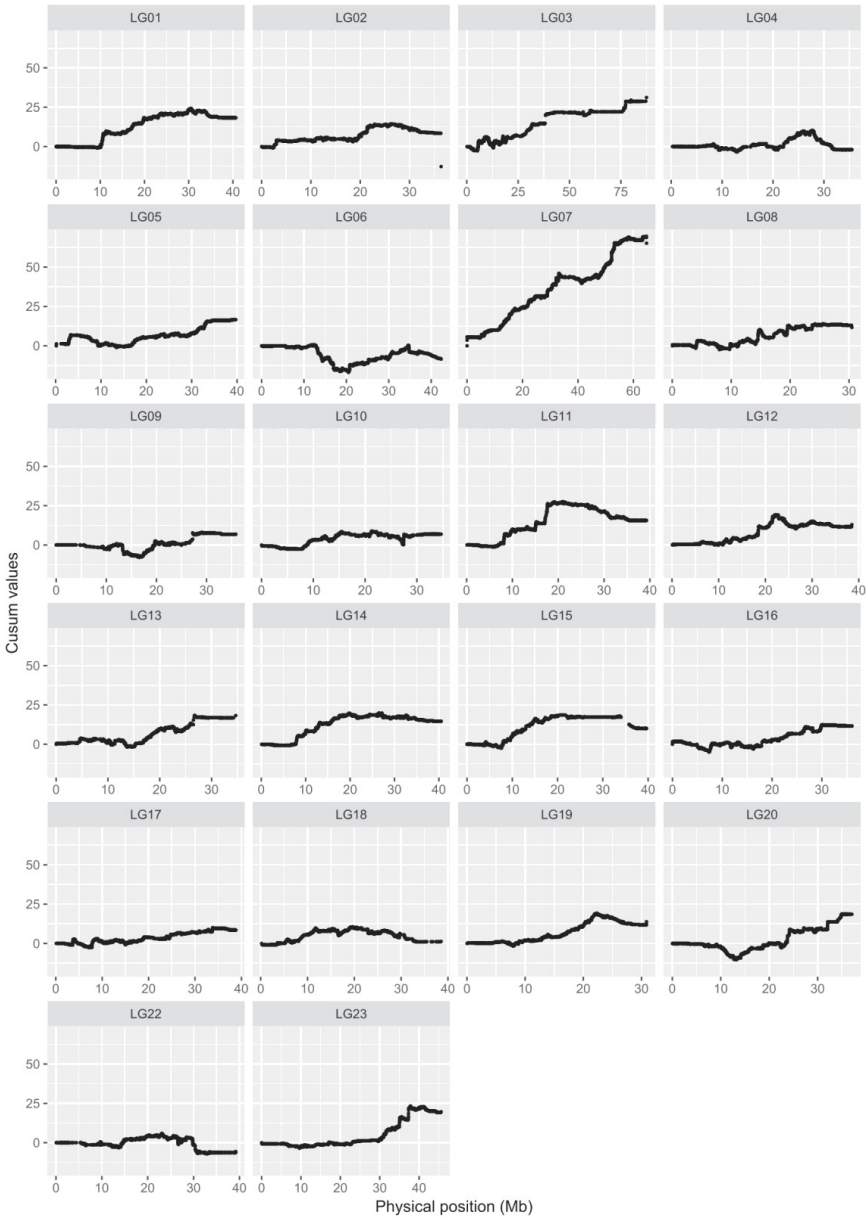
Supplementary Figure 2: The high-density consensus (sex-averaged) Build2 linkage map for Nile tilapia. The density is measured as cM/locus (higher the value, lower the number of markers in that locus)



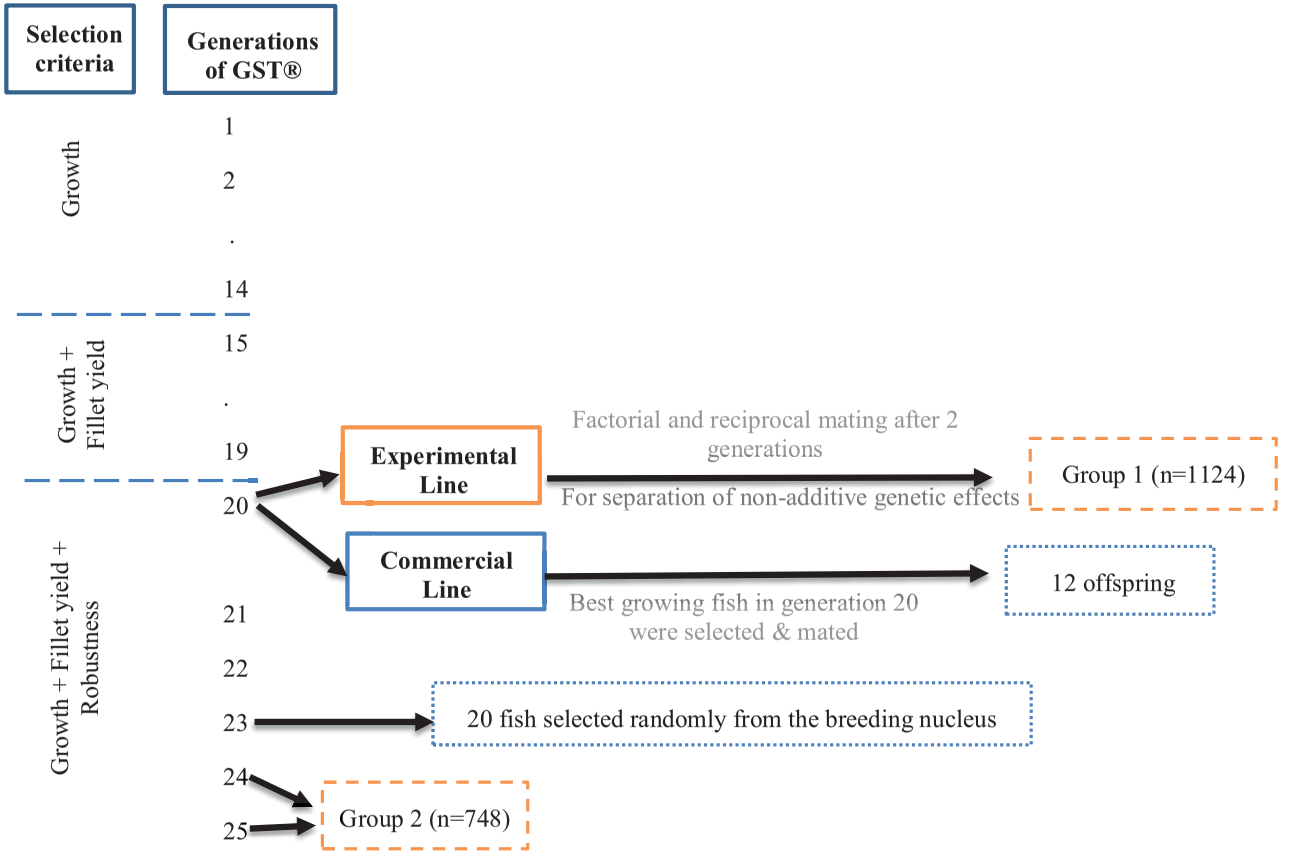
Supplementary Figure 3: The high-density female sex specific Build2 linkage map of Nile tilapia. The density is measured as cM/locus (higher the value, lower the number of markers in that locus)



Supplementary Figure 4: The high-density male sex specific Build2 linkage map of Nile tilapia. The density is measured as cM/locus (higher the value, lower the number of markers in that locus



Supplementary Figure 5: Cusum plots indicating the sex- patterned differentiations in the genome of tilapia. Cusum is a time series technique to emphasise shifts. Flat lines represent no difference between female and male recombination rate. Upward vertical lines represent more recombination rate in female LGs and downward vertical lines represent more recombination rate in male LGs.



Supplementary Figure 6: Selection criteria, sampling and genotyping design. Fish from Group 1 and Group 2 were used for linkage map construction. Group 1 (n=1124) comprised individuals collected following the branching of the 20th generation, and were factorially crossed against each other after 2 generations. Fish from Group 2 (n=748) were obtained from the 24th and 25th generations of GST®. 12 fish from commercial line and 20 fish from generation 23 were used for whole genome sequencing. Fish from group 1, Group 2 and 3119 fish from generations 20, 21 and 25 of breeding nucleus were used for array performance and validation.

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Manuscript

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Rajesh Joshi (corresponding author), Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway. +4767232739. rajesh.joshi@nmbu.no

John Woolliams, The Roslin Institute, Royal (Dick) School of Veterinary Studies, University of Edinburgh, Easter Bush, Midlothian, EH25 9RG, United Kingdom. And Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway. john.woolliams@roslin.ed.ac.uk

Theodorus Meuwissen, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway. +4767232739. theo.meuwissen@nmbu.no

Hans Magnus Gjøen, Department of Animal and Aquacultural Sciences, Norwegian University of Life Sciences, P.O. Box 5003, 1432 Ås, Norway. +4767232739. hans.magnus.gjoen@nmbu.no

Abstract

***Background**:*

The availability of both pedigree and genomic sources of information for animal breeding and genetics has created new challenges in understanding how best they may be utilized and how they may be interpreted. This study computed the variance components obtained using genomics and compared these to the variances obtained using pedigree in a population generated to estimate non-additive genetic variance. Further, the impact of assumptions concerning HWE on the component estimates was examined. The magnitude of inbreeding depression for important commercial traits in Nile tilapia was estimated for the first time, here using genomic data.

Results

The non-additive genetic variance in the Nile tilapia population was estimated from full-sib families and, where present, was found to be almost entirely additive by additive epistatic variance, although in pedigree studies this source is commonly assumed to arise from dominance. Further, substantial contributions ($P<0.05$) from non-additive genetic effects to the phenotypic variation of body depth (BD) and body weight at harvest (BWH) was found, with the estimates of the additive by additive epistatic ratio of 0.15 and 0.17 in the current breeding population using genomic data. In addition, the study found maternal variance ($P<0.05$) for BD, BWH, body length (BL) and fillet weight (FW) explaining approximately 10% of the observed phenotypic variance, comparable to the pedigree based estimates. The study also showed the detrimental effects of the inbreeding in the commercial traits of tilapia, and was estimated to be 1.08%, 0.91%, 0.37% and 0.34% decrease in the trait value with 1% increase in the individual homozygosity for FW, BWH, BD and BL, respectively. The

inbreeding depression and lack of dominance variance was consistent with an infinitesimal dominance model.

Conclusions:

Reciprocal recurrent selection is not required to exploit the non-additive variation in the biological traits as it will convert to additive genetic variation over time, although commercially this conclusion will depend on cost structures. However, the creation of maternal lines in Tilapia breeding schemes may be a possibility if this variation is found to be heritable.

Keywords: Nile tilapia, genomic selection, dominance, epistasis, maternal variance, non-additive genetic effects, reciprocal, heritability, inbreeding depression

Background

This paper is a part of an extensive study on the non-additive genetic effects in Nile tilapia and their utilization in tilapia breeding programs. The first part of this study [1] used the classical approach to partition the variance observed from a diallel mating design into additive, non-additive and maternal components using the pedigree information to generate the additive and dominance relationship matrixes. These variance components are inferred from the variances within and between full-sib families, where the latter is also decomposed among sires and among dams.

These pedigree based selection methods have been gradually supplemented with, or replaced by, genomic information in various livestock species [2], and even in some commercial aquaculture species [3]. With the possibility of improved accuracy and more detailed information from genomics [4], there has been a growing interest to try to quantify and utilize the non-additive source of phenotypic variation. This new technology has introduced new challenges to fully understand the results of these methods and their equivalence to the classical decompositions based on pedigree. The availability of genomic information in Nile tilapia [5] has offered the opportunity to close this gap in an important aquacultural species. The first aim of this paper is to compare the genetic variance components obtained from using either genomics or pedigree information to generate the appropriate relationship matrixes in a design generated to estimate non-additive variances.

The genomic BLUP (GBLUP) model builds a matrix of relationships between all individuals of a population based on genomic data, and BLUP uses these relations to partition the variance and predict the breeding values. The assumptions used to construct these relationship matrices have a direct effect on the accuracy of the results. There are different methods to construct the relationship matrices, most of

them differing in the scaling parameters [6–8], which makes it difficult to make comparisons. One method of comparison has been published by Legarra (2016) [9], where it is shown that the scaling of the relationship matrices to the same reference population is necessary when variance parameters have been obtained using different relationship matrices. In constructing relationship matrices, assumptions are often made on the presence of Hardy-Weinberg equilibrium (HWE), (e.g. in the use of Van Raden matrices [7] , as used by GCTA [10]), and on managing the linkage disequilibrium (LD) [11]. These assumptions influence the orthogonality of the estimates of the variance components and hence the validity and generality of their biological interpretation. Thus, the second aim of this paper is to examine the impact of assumption of HWE on the relationship matrices and the consequences for the estimation.

Inbreeding depression is a natural phenomenon that is widely assumed to be deleterious for traits of commercial importance and thus has serious practical implications [12–15]. It has greater impact in populations with smaller effective population size (N_e) than in those with higher N_e, due to purging of deleterious alleles [16,17] and this makes it a concern to breeders since N_e is often restricted. Genomic data allows a direct assessment of the extent of homozygosity and its variation rather than a reliance on changes predicted as a consequence of pedigree inbreeding. Consequently, utilisation of genomic data may contribute to the better design and preparedness of breeding programs. To date, the authors are unaware of estimates of inbreeding depression in Nile tilapia, even using the pedigree. Thus, the final aim of this paper is to quantify the effect of inbreeding depression for important commercial traits in Nile tilapia by the use of genomic data.

Methodology

Experimental design, phenotypes and genotypes:

The population used in this study and the experimental design have been previously described [1]. In short, the population was obtained from the reciprocal crossing of 2 parent groups, A and B, of Nile tilapia. Each parent group consisted of fullsibs only, although it was later found that one individual used as parent was half-sib. The matings were factorial so that each parent used, male or female, had offspring that were both full-sibs and half-sibs. Offspring obtained from these matings were reared in three batches and harvested over 8 different days after 6-7 months in the grow-out ponds. The fish were filleted by three filleters. The phenotypes recorded were body weight at harvest (BWH), body depth (BD), body length (BL), body thickness (BT), fillet weight (FW) and Fillet yield (FY). Phenotypes were obtained on a total of 2524 individuals, with 1318 and 1206 from each of the two reciprocal crosses, in altogether 155 full-sib families.

From these, 1882 Nile tilapia were genotyped (see Joshi et al. (2018) [5] for details) using the Onil50 SNP-array. The raw dataset contained 58,466 SNPs, which were analysed using the Best Practices Workflow with default settings (sample Dish QC ≥ 0.82, QC call rate ≥ 97; SNP call-rate cutoff ≥ 97) in the Axiom Analysis Suite software [18]. Ten samples were excluded based on low call rate. Then the SNPs were selected based on the informativeness, i.e. based on the formation of clusters and resolution. Only SNPs classified as PolyHighResolution (formation of three clusters with good resolution) and NoMinorHom (formation of two clusters with no samples of one homozygous genotype) were selected, and 43,014 SNPs were retained. The mean SNP call rate for these SNPs was 99.5 (range: 97-100). Finally, SNPs were filtered for minor allele frequency (MAF ≥ 0.05), and 39,927 SNPs (68.3% of the total genotyped

SNPs) were retained after all the quality control parameters had been applied. From the marker genotypes, the individual homozygosity was calculated as the proportion of homozygous loci per individual, and was incorporated into the models described below as a measure of directional dominance [19].

Out of the 1882 genotyped, 1119 individuals from 74 full-sib families with an average of 15.1 offspring per full-sib family (range 1 to 44; standard deviation = 11.2) had the phenotypic observation and were used for further analysis. The data structure is given in Table S1.1 of Supplementary 1 and the scatterplots and the phenotypic correlation of the phenotypes on these individuals are presented in Figure S1.1 of Supplementary 1. The descriptive statistics for the traits are given in Table S1.2 of Supplementary 1.

Statistical Analysis

ASReml-4 [20] was used to fit mixed linear models, using REML to estimate variance components and breeding values. Eight different univariate GBLUP models were tested and compared for the six traits described above. The basic model used was an animal model (A), which were gradually expanded to an ADME (model with additive, dominance, first order epistatic interactions and maternal effect) by adding each effect as random effects in a heuristic approach. This resulted in the following models:

A model: $\mathbf{y = X\beta + hb + Z_1a + e}$

AD model: $\mathbf{y = X\beta + hb + Z_1a + Z_2d + e}$

ADE model $\mathbf{y = X\beta + hb + Z_1a + Z_2d + Z_3e_{aa} + e}$

ADME model $\mathbf{y} = \mathbf{X\beta + hb} + \mathbf{Z_1a} + \mathbf{Z_2d} + \mathbf{Z_3e_{aa}} + \mathbf{Z_4m} + \mathbf{e}$

ADM model: $\mathbf{y = X\beta + hb + Z_1a + Z_2d + Z_4m + e}$

AM model: $\mathbf{y = X\beta + hb + Z_1a + Z_6m + e}$

AME model $\mathbf{y = X\beta + hb + Z_1a + Z_3e_{aa} + Z_4m + e}$

AE model: $\mathbf{y = X\beta + hb + Z_1a + Z_3e_{aa} + e}$

where, **y** is the vector of records; **β** is the vector of fixed effects that accounting for reciprocal cross (1 d.f.), batch (2 d.f.) and day of harvest (7 d.f.); **h** the vector of overall marker homozygosity for each individual, with b the inbreeding depression parameter; **a** is a vector of random additive genetic effects; **d** is vector of random dominance effects; $\mathbf{e_{aa}}$ is the vectors of first order additive x additive epistatic interactions; **m** is vector of maternal effects; **e** is a vector of random residual errors; and **X**, $\mathbf{Z_1}$, $\mathbf{Z_2}$, $\mathbf{Z_3}$ and $\mathbf{Z_4}$, are corresponding design matrices for fixed and random effects. For FW and FY, the fixed model also included filleter (2 d.f.). The (co)variance structures of the random effects are described below. Vectors **a, d,** $\mathbf{e_{aa}}$ and **e** had effects for each individual having genotypes; **m** for each maternal family.

The distributional assumptions for the random effects were multivariate normal, with mean zero and

$$\mathrm{Var}\begin{bmatrix} a \\ d \\ e_{aa} \\ m \\ e \end{bmatrix} = \begin{bmatrix} \boldsymbol{G}\sigma_A^2 & 0 & 0 & 0 & 0 \\ 0 & \boldsymbol{D}\sigma_D^2 & 0 & 0 & 0 \\ 0 & 0 & k(\boldsymbol{G}\#\boldsymbol{G})\sigma_{E_{aa}}^2 & 0 & 0 \\ 0 & 0 & 0 & \boldsymbol{I}\sigma_M^2 & 0 \\ 0 & 0 & 0 & 0 & \boldsymbol{I}\sigma_E^2 \end{bmatrix}$$

where σ^2_A, σ^2_D, σ^2_{Eaa}, σ^2_M and σ^2_E are additive genetic variance, dominance genetic variance, additive by additive epistatic variance, maternal variance and error variance respectively; **G** is the genomic relationship matrix; **D** is the dominance relationship matrix and **I** is an identity matrix of appropriate size. k(**G#G**) represents the additive by additive epistatic relationship matrix, where k is the scaling factor as described

below and # is the Hadamard product of the two matrices given by $(\mathbf{G\#G})_{ij} = g_{ij}^2$ for elements in the indices i and j.

The phenotypic variance was calculated as $\sigma^2_P = \sigma^2_A + \sigma^2_D + \sigma^2_{Eaa} + \sigma^2_M + \sigma^2_E$, and the estimated variance components were expressed relative to the total phenotypic variance (σ^2_P): additive heritability (h^2) = σ^2_A / σ^2_P, dominance ratio (d^2) = σ^2_D / σ^2_P and maternal ratio (m^2) = σ^2_M / σ^2_P. Broad sense heritability (H^2) was calculated as $(\sigma^2_A + \sigma^2_D + \sigma^2_{Eaa}) / \sigma^2_P$ and the terms not in a model were set to 0. The variances obtained were also scaled by $\overline{diag(\mathbf{V})} - \overline{\mathbf{V}}$ where **V** is their corresponding (co)variance matrix of size n and the bar denotes the mean value.

Genomic natural and orthogonal interactions (NOIA) and Hardy-Weinberg Equilibrium (HWE) approaches were used to calculate the **G**, **D** and k(**G#G**) following the methods of [21]. These approaches differ in two ways: (i) the contrasts between genotypes used to define dominance deviations, and (ii) the scaling factors used for the relationship matrices.

The NOIA approach relaxes the assumption of HWE in the population, under which the genomic relationship matrix ($\mathbf{G}$) is defined as:

$$\mathbf{G} = \frac{\boldsymbol{H_a H_a'}}{tr(\boldsymbol{H_a H_a'})/n}$$

where, $\mathbf{H_a}$ contains additive coefficients (h_a) having the dimension of *n* x *m*, with *n* = number of animals and *m* = number of SNPs. h_a is coded as:

$$h_a = \begin{cases} 2(1-p_A) \\ (1-2p_A) \\ -2p_A \end{cases} \quad \text{for genotypes} \begin{cases} AA \\ AB \\ BB \end{cases}$$

where, p_A is the frequency of allele *A*. For dominance deviations, NOIA uses the contrast that is orthogonal to h_a at each locus. Therefore, if p_{AA}, p_{AB} and p_{BB} are the allelic frequencies of the respective genotypes, the dominance relationship matrix (**D**) is defined as;

$$\mathbf{D} = \frac{\boldsymbol{H_d H_d'}}{tr(\boldsymbol{H_d H_d'})/n}$$

where, $\mathbf{H_d}$ contains dominance coefficients (h_d) defined for animal *i* and marker *j* by:

$$h_d = \begin{cases} -\frac{2p_{AB}\,p_{BB}}{p_{AA}+p_{BB}-(p_{AA}-p_{BB})^2} \\ \frac{4p_{AA}\,p_{BB}}{p_{AA}+p_{BB}-(p_{AA}-p_{BB})^2} \\ -\frac{2p_{AA}\,p_{AB}}{p_{AA}+p_{BB}-(p_{AA}-p_{BB})^2} \end{cases} \quad \text{for genotypes} \begin{cases} AA \\ AB \\ BB \end{cases}$$

The epistatic relationship matrices were then calculated from the Hadamard projects and scaled using the average of the diagonals. Therefore, the additive by additive epistatic relationship was calculated as:

$$\mathrm{k}(\mathbf{G\#G}) = \frac{\boldsymbol{G \# G}}{tr(\boldsymbol{G \# G})/n}$$

The HWE approach assumes that the population is under HWE equilibrium both in its scaling and in calculating the contrast for defining dominance deviations. If the locus is not in HWE the dominance contrast is not orthogonal to that for the additive effect, unlike in NOIA. The contrasts used to define the additive effects are unchanged but scaled assuming HWE, and the result is equivalent to method 1 of Raden [7]. So

$$\mathbf{G} = \frac{\boldsymbol{H_a H_a}'}{\sum 2p_i(1-p_i)}$$

where the sum in the denominator is over all *m* loci. The dominance relationship matrix was calculated as

$$\mathbf{D} = \frac{\mathbf{W_d W_d}'}{4\sum p_i^2 q_i^2}$$

where $\mathbf{W_d}$ contains elements w_d defined for animal *i* and marker *j*

$$w_d = \begin{cases} -2p_B^2 \\ 2p_A p_B \\ -2p_A^2 \end{cases} \quad \text{for genotypes} \begin{cases} AA \\ AB \\ BB \end{cases}$$

The scaling factor k for epistatic relationship matrices using the HWE approach was 1, so the additive by additive epistatic relationship matrix is simply the Hadamard product between the two matrices. The scatterplots for different relationship matrices are presented in Figure S1.3 and Figure S1.4 of Supplementary 1.

The software used to calculate the matrices did not accept missing genotypes. As described above, 0.4% of genotypes were missing and these were predicted using R code [22] by sampling from {0,1,2} with the probabilities for each given by observed probabilities for that SNP. The effect of this prediction was checked with GCTA [10] by constructing the GRMs including and excluding the imputed genotypes. The correlation of 1 between the additive and dominance relationships constructed using these two sets of genotypes suggest that there is no significant effect of prediction of the missing genotypes on our results as seen from the scatterplots of relationships in Figure S1.2 of Supplementary 1.

Comparison of Models

Likelihood ratio tests were used to measure the goodness of fit for the models. The critical values were corrected for boundary effects following [23]. The critical values are obtained from a mixture of χ^2 distributions with different degrees of freedom (d.f.) and were obtained for standard thresholds by iteration using R. The distributions of the likelihood under the null hypothesis of zero variances for 1, 2 and 3 components

were ½ I[0] + ½ χ^2_1, ¼ I[0] + ½ χ^2_1 + ¼ χ^2_2 and ⅛ I[0] +⅜ χ^2_1 + ⅜ χ^2_2+ ⅛ χ^2_3 where I[0] corresponds to a point mass of 1 at x=0.

Results

Genetic architecture

The six traits could be differentiated into three distinct groups based on their likelihood ratio tests (Table 1): BD and BWH showed evidence of significant maternal environmental effects and non-additive genetic effects in the form of additive by additive epistasis. BL and FW showed evidence of significant maternal environmental effects only; whereas BT and FY showed no evidence of neither maternal environmental nor additive by additive epistatic effects. None of the traits showed significant dominance variance. The assumption of HWE in the breeding population did not influence the goodness of fit for any of the model, because the log likelihood values were same.

Inbreeding depression

Detrimental effects of the genomic inbreeding were evident for all of these commercial traits, although of different magnitudes. BWH and FW were found to be more sensitive to inbreeding than the other traits, with nearly 1% decrease in the trait value per 1% increase in the individual homozygosity (Table 2). Traits BT and FY, the two traits with no evidence of non-additive genetic and maternal environmental effects, were found to be least sensitive.

Decomposition of variance components

Estimates of the variance components with the two approaches applied for all the models and traits are presented graphically in Figure 1. The summary table for the models selected based on likelihood ratio test are presented in Table 3, whereas the full table is available in Supplementary 2.

As expected, the simple A model gave the higher additive genetic variances, and the higher heritabilities across all the traits. Addition of dominance in the models had no effect on the estimated additive genetic variances, whereas including the additive by additive epistatic effect reduced the additive genetic variances markedly, except for BT and FY where there was no evidence (P>0.05) of epistasis. Inclusion of maternal environmental effect reduced the additive genetic variance compared to what was estimated with the simple A model, because without the maternal effect the additional variance associated with dams was interpreted as evidence of additive genetic effects. The maternal effect also reduced additive by additive epistatic variance in AE models. These reductions were again minimal for BT and FY. Similar results were obtained in both the NOIA and HWE assumption approaches. Hence, the numerical values are showed for the NOIA approach (scaled to the reference population [9]) , unless otherwise mentioned.

A model dependent variation in the estimation of additive variance was also reflected in the heritability estimates. For BT and FY, the two traits where the model of best fit was the simple A model, the heritabilities were least dependent on the models. For other traits, the difference among the models was up to 50%. For the best fit models, the estimates of the heritabilities were low to moderate, ranging from 0.08 ± 0.03 for BL to 0.19 ± 0.04 for FY (Table 4).

For BD and BWH, the traits for which the best fit model included additive by additive epistatic effect, the additive by additive epistatic ratio (e_{aa}^2) was 0.15 ± 0.09 and 0.17 ± 0.10 (Table 4). Additive by additive epistasis was found to be 48% and 63% of the total genetic variance for BD and BWH, respectively, but with huge standard errors. Various other papers with genomic epistatic models also report huge epistatic components [21,24,25] with corresponding huge standard errors. Large differences

between the individuals (Figure 2a) and the full-sib families (Figure 2b) were observed for the additive by additive epistatic effects.

For the four traits where the model of best fit included maternal environmental effect, maternal ratio was found to be around 0.08±0.04 to 0.09± 0.06. As expected, this variance ratio was not affected by the two approaches or the models used.

Discussion

Interpretation of variance within the full-sib family

A major finding of this study is that the use of genomic relationship matrixes identified the source of non-additive genetic variance as being almost entirely additive by additive epistatic variance. The primary source of non-additive variance is commonly assumed to be dominance in pedigree based analyses [1,26,27], but this assumption seems to be misleading with the estimates of dominance variance being negligible. In this study, the information for estimating non-additive variance comes from the variance within full-sib families (see Supplementary Information 3), and in the presence of dominance and epistasis, the additional variance in full-sib families, above the additive variance provided by the sire and dam, is $¼\ \sigma^2_D + ⅛\ \sigma^2_{Eaa} + ⅛\ \sigma^2_{Ead} + ⅛\ \sigma^2_{Edd}$, where σ^2_D, σ^2_{Eaa}, σ^2_{Ead} and σ^2_{Edd} are dominance, additive by additive, additive by dominance and dominance by dominance epistatic variances [28]. Under an infinitesimal model with both additive and dominance effects, with the increase in the number of loci, either the dominance variance tends towards zero or the inbreeding depression tends towards infinity [28,29]. Thus, dominance may be present, but the genomic approach is showing this component behaves infinitesimal, with $\sigma^2_{D,}$ σ^2_{Ead} and σ^2_{Edd} bound to zero.

Comparison with pedigree approach

This study adds a new dimension to our previous paper [1] and to the best of our knowledge, these are the first published results using genomic information in tilapia. The availability of the genomic data in populations will inevitably lead to comparisons of genomic- and pedigree-based heritabilities but these are not straightforward. Some publications argue that pedigree-based methods overestimate heritabilities [30–32], while some suggest the reverse [33–36], and other that the heritabilities are similar [37].

However, few studies recognize that the variance parameters obtained even in basic additive models are not estimates of the variance for the same populations, and therefore the simple comparison of parameters can be meaningless. For pedigree-based analyses the estimated variance parameter refers to the base population of the pedigree (a subset of **A**), and for genomic-based analyses it can be viewed as the genetic variance in the population is defined by the whole **G** *and* assuming all the markers are in HWE. Many papers compare the uncorrected values and are therefore uninformative, as large part of difference in variances can be accounted for by such differences [9,21]. To overcome the problem of comparability, the variance parameters from NOIA and HWE approaches were used to estimate the genetic variance in the entire population of this study [9] with marker genotypes as observed, equivalent to scaling the variance component estimates by $\overline{diag(\mathbf{V})} - \overline{\mathbf{V}}$, where **V** is the relevant relationship matrix**.**

In this study, where the models go beyond the additive components, there are additional reasons why components may differ. In the tilapia population studied here, the additive variance, when dominance is assumed to be the source of non-additive variation, gives a qualitatively different estimate to that obtained if additive epistasis is

assumed (see Supplementary 3). Therefore, differences should be expected between the current study and [1]. A further issue with this study was that the data used was only a subset of the data used for [1], although Figure S1.5 of Supplementary 1 shows the sampling does not deviate far from random sampling expectations. This issue was overcome by repeating the pedigree analyses using only the phenotypes included in this study (see Table S1.4 in Supplementary 1).

The outcome from objective comparisons of the pedigree- and genomic analyses showed a qualitatively similar pattern of contributing sources of variance for all 6 traits insofar as additive, maternal and non-additive variances. The impact on the genetic evaluations was also similar (see supplementary 5). Some small differences were observed: for example, although the quantitative outcomes for the maternal ratio were similar, the qualitative outcomes for statistical significance thresholds showed differences for BT and BL. The evidence of non-additive genetic effects was found for the same traits (BD, BWH) irrespective of the type of relationships used. However, as mentioned above, critically, the genomics identified the source of non-additivity as additive by additive epistasis rather than dominance.

Using the basic model A with no other sources of variance, uncorrected pedigree heritabilities were as much as 60% greater than genomic heritabilities, but this discrepancy decreased after correction following [9] and were broadly comparable (relative to their standard errors; see Table S1.4 of Supplementary 1), particularly for FY and BT. Both these traits favored simple additive models, but for the other traits, the genetic architecture of the traits favored maternal and non-additive genetic effects. In the basic model A, using pedigree, the dam information is absorbed into estimating additive variance; in contrast to the genomic model, where it is the genotypes of the dam and its offspring that contribute information on the heritabilities, so the dam

363 variance is no longer (wrongly) absorbed into the additive variance. Hence the
364 pedigree-based heritabilities are higher for traits with maternal variance, as a
365 consequence of the wrong model. These comparison shows the difference in
366 sensitivity between the pedigree model and the genomic model to mis-specification of
367 sources of variance in both the genetic architecture and the environmental model.

368 ## Impact of approaches used

369 GBLUP uses GRMs, and the assumptions in the construction of these GRMs can have
370 a direct effect. Several approaches (e.g. Van Raden matrices [7]) assume Hardy
371 Weinberg equilibrium when scaling the relationship matrices, which is relaxed in the
372 NOIA approach. The use of these genomic approaches showed no difference to the
373 qualitative parameter like genetic architecture of the trait, but quantitative differences
374 were observed for the sources of variation, with additive by additive epistatic ratio
375 (e_{aa}^2) being inflated by ca. 20 % and 18%, and heritability (h^2) by 6% and 10% for the
376 traits BD and BWH respectively (Table 3); this has also been observed in other studies
377 [21]. However, the transformation of these variance components of both NOIA and
378 HWE assumption approach on a similar scale based on the relationship matrices yield
379 identical additive genetic variance and additive by additive epistatic variance (Table
380 4). But, the heritabilities and epistatic ratios were not similar, although a decrease in
381 the differences could be seen, due to the differences in the residual variance in both
382 approaches. As a consequence of the absence of dominance variance, the differences
383 between the NOIA and HWE collapse into differences in the scaling of the relationship
384 matrices as the contrasts used to construct the matrices were identical.

385 The NOIA and HWE approaches are statistical models in that they partition the
386 variance observed in a population and use these parameters to estimate breeding
387 values and dominance deviations [21]. As such, these estimates depend on the allele

388 frequencies in the particular population and the structure of the population, which will
389 influence the genotypic frequencies. A distinction needs to be made between the
390 magnitudes of the variance components in the total genetic variance and the effects
391 estimated using them on the one hand, and the ubiquity of the same phenomena in
392 genotypic models (sometimes called biological models) on the other hand [38,39]. For
393 example, the genotypes at a single locus may show complete dominance, but have a
394 negligible dominance deviation, because the superior homozygote is very rare in the
395 population. Although the NOIA approach removes limitations of HWE, there are major
396 barriers to it moving towards the building of genotypic models. Firstly, it does not
397 remove the impact of LD on estimates of the effects, and more seriously, the genotypic
398 models are meaningful only if constructed with the causal variants and not with
399 anonymous markers.

400 ## Inbreeding depression

401 Absence of dominance variance does not necessarily mean the absence of inbreeding
402 depression when the genetic architecture approaches the infinitesimal model, and
403 evidence was found for depression in precisely the same four traits for which the basic
404 A model was rejected. To the authors' knowledge, these estimates are the first for the
405 commercial traits in Nile tilapia. Most of the quantification has been done using
406 pedigree information in other aquaculture species (e.g. [40–42]), and a handful using
407 genomics [43]. In the absence of genomics, this information was not observable
408 because of the near identical inbreeding coefficients of the study population. The
409 majority of the traits clearly show the signal of inbreeding depression and ignoring this
410 term leaves the estimates of the variance components and predictions of offspring
411 merit open to bias.

Utilisation of the additive by additive epistatic effects

The magnitude of the epistatic variance is argued [28,45] to be much smaller than the additive genetic variance in elite commercial populations for the following reasons: (i) all the interaction effects contribute to the additive variance; and (ii) additive variance is the function of heterozygosity at individual loci, whereas the epistatic variance is the function of the products of heterozygosity at multiple loci, which is generally low because of the U-shaped distributions in allele frequency that are typically observed. In the long run, additive by additive epistatic variance is expected to be exploited indirectly as it is converted to the additive genetic variance due to random drift and selection; hence this form of variance affects the medium and long-term selection response indirectly [44], and in itself does not argue for re-structuring towards reciprocal recurrent selection. On the one hand, while it has been argued that the epistatic effects should be included in the estimation of genetic parameters, since additive genetic variance already contains a fraction of the epistatic variances in the presence of the epistasis, it is likely to be unnecessary for selection decisions [24,28,45]. On the other hand, the huge epistatic ratio, predicting large differences among individuals in the population (Figure 2), prompts the question of whether this source can be used in the Nile tilapia breeding program in some way. This is particularly unattractive to breeders when depending on the interactions of anonymous loci. Therefore, more direct exploitation of epistasis in the future in Nile tilapia breeding program will depend on finding out the causal variants showing large epistatic interactions [46,47] for different traits. However, this will require a substantial resource to achieve, probably an order of magnitude greater than for identifying the additive effects of causal variants.

Conclusion

Thus, this study has found that the non-additive genetic variance in the Nile tilapia population was almost entirely additive by additive epistatic variance using genomic relationship matrixes, whereas this is commonly assumed to be dominance using pedigree based relationship matrixes. The inbreeding depression and lack of dominance variance was consistent with an infinitesimal dominance model. Reciprocal recurrent selection is not required to exploit the non-additive variation in the biological traits as it will convert to additive genetic variation over time, although commercially this conclusion will depend on cost structures. However, the creation of maternal lines in Tilapia breeding schemes may be a possibility if this variation is found to be heritable.

List of abbreviations

Acronym	Full Form
BD	Body Depth
BL	Body Length
BT	Body Thickness
BWH	Body Weight at Harvest
d.f.	degrees of freedom
FW	Fillet Weight
FY	Fillet Yield
GRMs	Genomic relationship matrices
HWE	Hardy-Weinberg Equilibrium
LD	Linkage Disequilibrium
NOIA	Natural and orthogonal interactions

Declarations

Ethics approval and consent to participate: Not applicable

Consent for publication: Not applicable

Availability of data and material

The genotype data used in the study are from commercial family material. This information may be made available to non-competitive interests under conditions specified in a Data Transfer Agreement. Requests to access these datasets should be directed to Alejandro Tola Alvarez: alex@genomar.com.

Competing interests

The authors declare that they have no competing interests

Funding: Not applicable

Authors' contributions

HMG conceived and designed the study, RJ did the statistical analysis, JAW contributed to this analysis, and all authors contributed to the discussion of the results and writing of the paper.

Acknowledgements

We would like to thank GenoMar AS for providing the data, in particular Anders Skaarud and Alejandro Tola Alvarez. ASReml analysis was performed in the Abel Cluster, Oslo.

Authors' information (optional): Not applicable

Supplementary files

Supplementary 1: Data Structure and the relationship matrices

Supplementary 2: Full table showing the decomposition of variances for all models, traits and approaches, available at https://drive.google.com/file/d/1pXI1WtmcTxckTdM8LOfOCW94jPKcgFua/view?usp=sharing

Supplementary 3: Assumptions on the nature of non-additive genetic variance and the impact on estimates of additive genetic variance

Supplementary 4: Impact of inbreeding depression on models

Supplementary 5: Impact on the genetic evaluation

References

1. Joshi R, Woolliams J, Meuwissen T, Gjøen H. Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. Heredity (Edinb). [Internet]. Nature Publishing Group; 2018 [cited 2018 Jan 16];1. Available from: http://www.nature.com/articles/s41437-017-0046-x

2. Meuwissen T, Hayes B, Goddard M. Genomic selection: A paradigm shift in animal breeding. Anim. Front. American Society of Animal Science; 2016;6:6–14.

3. Hosoya S, Kikuchi K, Nagashima H, Onodera J, Sugimoto K, Satoh K, et al. Genomic Selection in Aquaculture. Bull. Jap. Fish. Res. Edu. Agen. No. 2017;45:35–9.

4. Gianola D, Fernando RL, Stella A. Genomic assisted prediction of genetic value with semi-parametric procedures. Genetics. Genetics Soc America; 2006;

5. Joshi R, Arnyasi M, Lien S, Gjoen HM, Alvarez AT, Kent M. Development and validation of 58K SNP-array and high-density linkage map in Nile tilapia (O. niloticus). bioRxiv. Cold Spring Harbor Laboratory; 2018;322826.

6. Speed D, Balding DJ. Relatedness in the post-genomic era: is it still useful? Nat. Rev. Genet. Nature Publishing Group; 2015;16:33.

7. VanRaden PM. Efficient Methods to Compute Genomic Predictions. J Dairy Sci. 2008;91:4414–23.

8. Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, et al. Common SNPs explain a large proportion of the heritability for human height. Nat Genet [Internet]. 2010;42:565–9. Available from: http://dx.doi.org/10.1038/ng.608

9. Legarra A. Comparing estimates of genetic variance across different relationship models. Theor. Popul. Biol. Elsevier; 2016;107:26–30.

10. Yang J, Lee SH, Goddard ME, Visscher PM. GCTA: a tool for genome-wide complex trait analysis. Am. J. Hum. Genet. Elsevier; 2011;88:76–82.

11. Speed D, Cai N, Johnson MR, Nejentsev S, Balding DJ, Consortium U. Reevaluation of SNP heritability in complex human traits. Nat. Genet. Nature Publishing Group; 2017;49:986.

12. Fessehaye Y, Bovenhuis H, Rezk MA, Crooijmans R, van Arendonk JAM, Komen H. Effects of relatedness and inbreeding on reproductive success of Nile tilapia (Oreochromis niloticus). Aquaculture. Elsevier; 2009;294:180–6.

13. Christensen K, Jensen P, Jørgensen JN. A note on effect of inbreeding on production traits in pigs. Anim. Sci. Cambridge University Press; 1994;58:298–300.

14. Smith LA, Cassell BG, Pearson RE. The effects of inbreeding on the lifetime performance of dairy cattle. J. Dairy Sci. Elsevier; 1998;81:2729–37.

15. Bjelland DW, Weigel KA, Vukasinovic N, Nkrumah JD. Evaluation of inbreeding depression in Holstein cattle using whole-genome SNP markers and alternative measures of genomic inbreeding. J. Dairy Sci. Elsevier; 2013;96:4697–706.

16. Meuwissen THE, Woolliams JA. Effective sizes of livestock populations to prevent a decline in fitness. Theor. Appl. Genet. Springer; 1994;89:1019–26.

17. Kristensen TN, Sørensen AC. Inbreeding–lessons from animal breeding, evolutionary biology and conservation genetics. Anim. Sci. Cambridge University Press; 2005;80:121–33.

18. Thermo Fisher Scientific Inc. Axiom™Analysis Suite (AxAS) v4.0 USER GUIDE [Internet]. 2018. Available from: https://downloads.thermofisher.com/Affymetrix_Softwares/Axiom_Analysis_Suite_AxAS_v4.0_User_Guide.pdf

19. Xiang T, Christensen OF, Vitezica ZG, Legarra A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. Genet. Sel. Evol. [Internet]. BioMed Central; 2016 [cited 2017 May 15];48:92. Available from: http://www.ncbi.nlm.nih.gov/pubmed/27887565

20. Gilmour A, Thompson R. ASReml 4 Australasian Statistics Conference Port Lincoln 2014. 2014;

21. Vitezica ZG, Legarra A, Toro MA, Varona L. Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. Genetics [Internet]. 2017 [cited 2018 Jan 29];206:1297–307. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28522540

22. Mesfingo. R-code-for-missing-genotype-imputation [Internet]. 2016 [cited 2018 Jan 3]. Available from: https://github.com/Mesfingo/R-code-for-missing-genotype-imputation

23. Visscher PM. A note on the asymptotic distribution of likelihood ratio tests to test variance components. Twin Res. Hum. Genet. [Internet]. Cambridge University Press; 2006;9:490–5. Available from: https://doi.org/10.1375/twin.9.4.490

24. Raidan FSS, Porto-Neto LR, Li Y, Lehnert SA, Vitezica ZG, Reverter A. Evaluation of non-additive effects in yearling weight of tropical beef cattle. J. Anim. Sci. 2018;

25. Piaskowski J, Hardner C, Cai L, Zhao Y, Iezzoni A, Peace C. Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. BMC Genet. BioMed Central; 2018;19:23.

26. Shaw FH, Woolliams JA. Variance component analysis of skin and weight data for sheep subjected to rapid inbreeding. Genet. Sel. Evol. BioMed Central; 1999;31:1.

27. Hill WG. On estimation of genetic variance within families using genome-wide identity-by-descent sharing. Genet. Sel. Evol. BioMed Central; 2013;45:32.

28. Falconer DS, Mackay TF, Frankham R. Introduction to Quantitative Genetics (4th edn). Trends Genet. 1996;12:280.

29. Toro MA, Mäki-Tanila A. Some intriguing questions on Fisher's ideas about dominance. J. Anim. Breed. Genet. Wiley Online Library; 2018;135:149–50.

30. Haile-Mariam M, Nieuwhof GJ, Beard KT, Konstatinov K V, Hayes BJ. Comparison of heritabilities of dairy traits in Australian Holstein-Friesian cattle from genomic and pedigree data and implications for genomic evaluations. J. Anim.

564 Breed. Genet. Wiley Online Library; 2013;130:20–31.

565 31. Bérénos C, Ellis PA, Pilkington JG, Pemberton JM. Estimating quantitative
566 genetic parameters in wild populations: a comparison of pedigree and genomic
567 approaches. Mol. Ecol. Wiley Online Library; 2014;23:3434–51.

568 32. Veerkamp RF, Mulder HA, Thompson R, Calus MPL. Genomic and pedigree-
569 based genetic parameters for scarcely recorded traits when some animals are
570 genotyped. J. Dairy Sci. [Internet]. 2011;94:4189–97. Available from:
571 http://www.sciencedirect.com/science/article/pii/S002203021100422X

572 33. Tsai H-Y, Hamilton A, Tinch AE, Guy DR, Gharbi K, Stear MJ, et al. Genome
573 wide association and genomic prediction for growth traits in juvenile farmed Atlantic
574 salmon using a high density SNP array. BMC Genomics. BioMed Central;
575 2015;16:969.

576 34. Bangera R, Correa K, Lhorente JP, Figueroa R, Yáñez JM. Genomic predictions
577 can accelerate selection for resistance against Piscirickettsia salmonis in Atlantic
578 salmon (Salmo salar). BMC Genomics. BioMed Central; 2017;18:121.

579 35. Tan B, Grattapaglia D, Wu HX, Ingvarsson PK. Genomic relationships reveal
580 significant dominance effects for growth in hybrid Eucalyptus. Plant Sci. Elsevier;
581 2018;267:84–93.

582 36. Zhang YD, Johnston DJ, Bolormaa S, Hawken RJ, Tier B. Genomic selection for
583 female reproduction in Australian tropically adapted beef cattle. Anim. Prod. Sci.
584 CSIRO; 2014;54:16–24.

585 37. Vitezica ZG, Varona L, Elsen J-M, Misztal I, Herring W, Legarra A. Genomic
586 BLUP including additive and dominant variation in purebreds and F1 crossbreds,
587 with an application in pigs. Genet. Sel. Evol. [Internet]. BioMed Central; 2016 [cited
588 2016 Feb 2];48:6. Available from:
589 http://gsejournal.biomedcentral.com/articles/10.1186/s12711-016-0185-1

590 38. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and
591 covariance of individuals within the genomic selection scope. Genetics [Internet].
592 2013 [cited 2016 Jul 8];195:1223–30. Available from:
593 http://www.ncbi.nlm.nih.gov/pubmed/24121775

594 39. Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive
595 genetic variance for complex traits. PLoS Genet. Public Library of Science;
596 2008;4:e1000008.

597 40. Rye M, Mao ILL. Nonadditive genetic effects and inbreeding depression for body
598 weight in Atlantic salmon (Salmo salar L.). Livest. Prod. Sci. [Internet]. Elsevier; 1998
599 [cited 2018 Sep 4];57:15–22. Available from:
600 https://www.sciencedirect.com/science/article/pii/S0301622698001651

601 41. Pante MJR, Gjerde B, McMillan I. Effect of inbreeding on body weight at harvest
602 in rainbow trout, Oncorhynchus mykiss. Aquaculture [Internet]. Elsevier; 2001 [cited
603 2018 Sep 4];192:201–11. Available from:
604 https://www.sciencedirect.com/science/article/pii/S0044848600004671

605 42. Neira R, Díaz NF, Gall GAE, Gallardo JA, Lhorente JP, Manterola R. Genetic
606 improvement in Coho salmon (Oncorhynchus kisutch). I: Selection response and

607 inbreeding depression on harvest weight. Aquaculture [Internet]. Elsevier; 2006
608 [cited 2018 Sep 4];257:9–17. Available from:
609 https://www.sciencedirect.com/science/article/pii/S0044848606001839

610 43. Hu G, Wang C, Da Y. Genomic heritability estimation for the early life-history
611 transition related to propensity to migrate in wild rainbow and steelhead trout
612 populations. Ecol. Evol. [Internet]. 2014 [cited 2015 Oct 16];4:1381–8. Available
613 from: http://doi.wiley.com/10.1002/ece3.1038

614 44. Hill WG. “Conversion” of epistatic into additive genetic variance in finite
615 populations and possible impact on long-term selection response. J. Anim. Breed.
616 Genet. Wiley Online Library; 2017;134:196–201.

617 45. Mäki-Tanila A, Hill WG. Contribution of gene–gene interaction to genetic
618 variation and its utilisation by selection. Proc. 10th World Congr. Genet. Appl. to
619 Livest. Prod. 2014.

620 46. Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L. Epistasis and the
621 release of genetic variation during long-term selection. Nat. Genet. Nature Publishing
622 Group; 2006;38:418.

623 47. Große-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, Jüngst H, et
624 al. Epistatic QTL pairs associated with meat quality and carcass composition traits in
625 a porcine Duroc× Pietrain population. Genet. Sel. Evol. BioMed Central; 2010;42:39.

626

627

List of tables

Table 1: Log likelihood values with significance levels for different models for the six traits. The significance level for the likelihood ratio tests are expressed relative to the full model ADME. The critical values for Type 1 errors of 0.05, 0.01 and 0.001 were: for 1 d.f., 2.71, 5.42 and 9.55, respectively; for 2 d.f., 4.24, 7.29 and 11.77; and for 3 d.f. 5.44, 8.75 and13.48 respectively. The statistical significance is labelled as '*', '**' and '***' for P<0.05, P<0.01 and P<0.001, respectively.

Models	d.f.	BD	BL	BT	BWH	FW	FY
ADME		-43.48	-191.28	-1.78	-31.51	-69.90	-68.55
ADE	**1**	-46.55**	-195.75**	-2.25	-35.82**	-74.74***	-69.10
ADM	**1**	-45.14*	-192.02	-2.34	-33.40*	-70.40	-68.65
AME	**1**	-43.48	-191.28	-1.78	-31.51	-69.90	-68.55
AD	**2**	-49.29**	-197.99***	-3.04	-39.29***	-76.05***	-69.25
AE	**2**	-46.55*	-195.75**	-2.25	-35.82**	-74.74**	-69.10
AM	**2**	-45.15	-192.02	-2.40	-33.40	-70.40	-68.65
A	**3**	-49.29**	-197.99**	-3.06	-39.29***	-76.05**	-69.25

637 **Table 2:** Inbreeding depression for the commercial traits in Nile tilapia. "b" is the
638 regression coefficient of trait on individual homozygosity, and D is the percentage
639 decrease in the trait value per 1% increase in the individual homozygosity due to
640 inbreeding depression. Standard errors are presented inside the parenthesis (). **
641 indicates p values 0.001 - 0.01 and * indicates p values 0.01 - 0.05 for significant
642 values.

	BD	BWH	BL	FW	BT	FY
b	-3.27** (1.19)	-371** (137)	-7.57* (2.95)	-156** (56)	-7.08 (5.05)	-6.90 (4.93)
D	0.37	0.91	0.34	1.08	0.17	0.21

643

644

Table 3: Components and their ratios with phenotypic variance for the models of best fit for different traits. Standard errors are presented in parentheses. The ratios are: narrow heritability h^2, broad heritability H^2, maternal ratio m2 and epistatic ratio e_{aa}^2

Trait	Model	σ^2_A	σ^2_{Eaa}	σ^2_m	σ^2_e	σ^2_p	h^2	H^2	m^2	e_{aa}^2
						NOIA				
BD	AME	0.086 (0.024)	0.080 (0.049)	0.047 (0.032)	0.328 (0.044)	0.541 (0.039)	0.158 (0.042)	0.307 (0.090)	0.087 (0.055)	0.148 (0.091)
BWH	AME	699 (268)	1183 (680)	635 (418)	4540 (618)	7059 (498)	0.099 (0.037)	0.266 (0.093)	0.090 (0.054)	0.167 (0.096)
BL	AM	0.284 (0.107)		0.257 (0.162)	2.803 (0.136)	3.345 (0.209)	0.085 (0.031)		0.076 (0.045)	
FW	AM	118 (42)		99 (63)	1009 (50)	1227 (79)	0.096 (0.033)		0.080 (0.047)	
BT	A	1.695 (0.441)			8.015 (0.411)	9.710 (0.458)	0.174 (0.041)			
FY	A	1.758 (0.406)			7.461 (0.378)	9.220 (0.435)	0.190 (0.039)			
						HWE				
BD	AME	0.097 (0.027)	0.102 (0.063)	0.047 (0.032)	0.326 (0.045)	0.573 (0.042)	0.169 (0.046)	0.348 (0.1)	0.082 (0.053)	0.178 (0.106)
BWH	AME	791 (303)	1504 (864)	635 (418)	4520 (626)	7450 (544)	0.106 (0.04)	0.308 (0.104)	0.085 (0.051)	0.201 (0.111)
BL	AM	0.321 (0.120)		0.257 (0.162)	2.801 (0.136)	3.380 (0.213)	0.095 (0.034)		0.076 (0.044)	
FW	AM	133 (47)		99 (63)	1009 (50)	1241 (81)	0.107 (0.036)		0.079 (0.047)	
BT	A	1.915 (0.498)			8.004 (0.413)	9.92 (0.492)	0.193 (0.044)			
FY	A	1.987 (0.459)			7.450 (0.379)	9.437 (0.467)	0.210 (0.042)			

Table 4: Corrected heritabilities, ratio and variances for the models of best fit for different traits and approaches. The variances and ratios were corrected by (Mean (leading diagonal) – Mean) of the the corresponding relationship matrices as per Legarra (2016). Standard errors are presented in parenthesis.

Traits	HWE				NOIA			
	σ^2_A	σ^2_{Eaa}	h^2	$e_{aa}{}^2$	σ^2_A	σ^2_{Eaa}	h^2	$e_{aa}{}^2$
BD	0.086 (0.024)	0.080 (0.049)	0.150 (0.041)	0.139 (0.083)	0.086 (0.024)	0.080 (0.049)	0.159 (0.043)	0.147 (0.091)
BWH	698.774 (267.730)	1169.547 (672.154)	0.094 (0.035)	0.157 (0.0867)	698.772 (267.729)	1169.539 (672.149)	0.099 (0.037)	0.166 (0.095)
BL	0.285 (0.107)		0.084 (0.030)		0.284 (0.107)		0.085 (0.031)	
FW	117.948 (41.825548)		0.095 (0.0324407)		117.948 (41.825)		0.096 (0.033)	
BT	1.694 (0.441)		0.171 (0.039)		1.694 (0.441)		0.174 (0.041)	
FY	1.757 (0.406)		0.186 (0.038)		1.758 (0.406)		0.191 (0.039)	

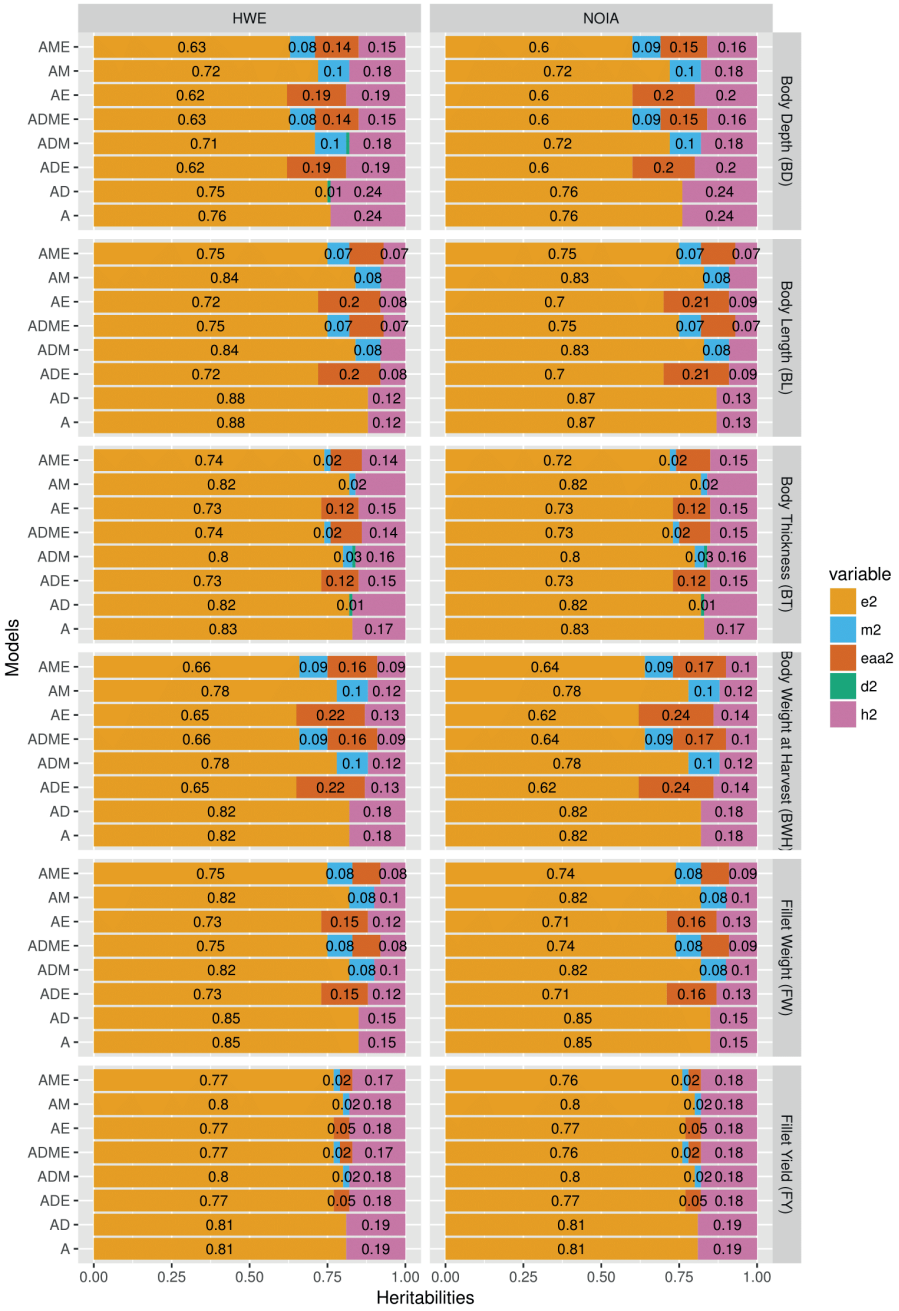


Figure 1: Decomposition of the phenotypic variance into different components using NOIA and HWE assumption approaches for the six traits. The ratios are: h^2 is additive; d^2 is dominance; e_{aa}^2 is additive by additive epistatic; m^2 is maternal; and e^2 is residual.

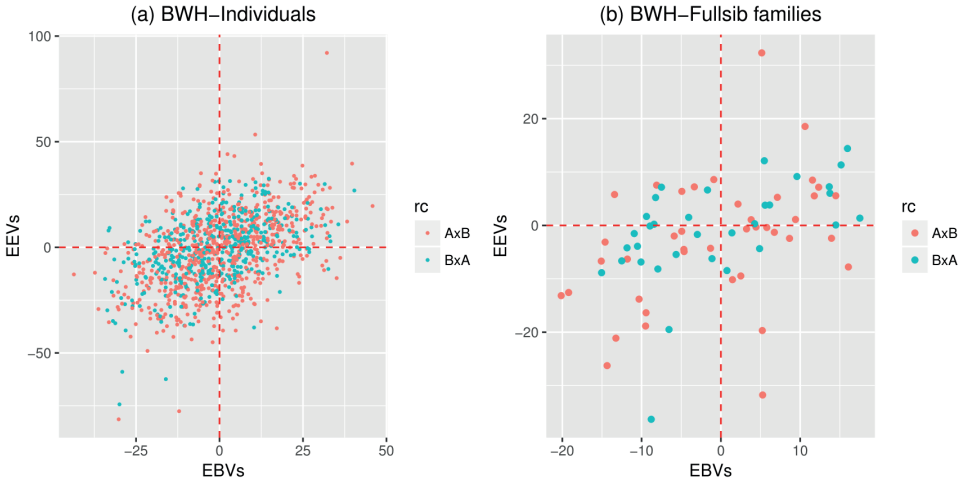


Figure 2: Scatterplot of estimated breeding values (EBVs) and epistatic (additive by additive) values (EEVs) for the trait BWH using NOIA approach (a) shows the scatterplot for all the individuals (b) shows the scatterplot for the mean values for different full-sib families. Please note that the values for x-axis and y-axis are different for both plots. The color of the dots in the scatterplot represents the types of reciprocal cross (rc): AxB and BxA.

667

Supplementary 1: Data structure and the relationship matrices

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Table S1.1: Number of animals genotyped in different full-sib families. The row with P and G denotes the total number of phenotyped (coded black) and genotyped animals (coded bold blue) respectively from each full-sib family. Parents are coded from 1 to 18 in the header rows and columns. Empty cells means no phenotypes or genotypes were available for that full-sib family.

		1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	16	17	18
1	**P**				5	9		30		5	5	26	26		34		15	35	
	G																		
2	**P**	30	20	18			11		33					9		3			16
	G	**30**	**20**	**18**			**11**							**8**					
3	**P**	29	18	25			6		36					8		13			36
	G																		
4	**P**	13	7	19			6		23					11		6			14
	G																		
5	**P**							6		8	7	20	12		25		10	27	
	G							**6**		**8**	**7**	**20**			**21**		**10**	**27**	
6	**P**				3	6		46		8	4	39	44		59		17	54	
	G							**41**		**8**	**4**	**39**			**38**		**17**	**44**	
7	**P**	32	9	17			8		32					7		2			30
	G	**32**	**9**	**17**			**8**							**7**					
8	**P**				5	3		24		9	3	13	13		19		8	16	
	G							**25**		**8**	**3**	**13**			**10**		**8**	**16**	
9	**P**	29	16	26			9		37					11		3			27
	G	**29**	**15**	**26**			**9**							**11**					
10	**P**	35	27	16			8		38					17		6			30
	G																		
11	**P**				1	2		13		2	1	8	13		22		8	21	
	G																		
12	**P**				4	2		36		4	3	10	30		47		10	52	
	G							**36**		**4**	**3**	**10**			**34**		**8**	**43**	
13	**P**	19	3	14			3		19					7		1			10
	G	**19**	**3**	**14**			**3**							**7**					
14	**P**				2	6		14				16	13		15		12	45	
	G							**14**				**16**			**9**		**12**	**43**	
15	**P**	16	15	22			6		28					12		4			29
	G	**16**	**15**	**20**			**6**		**1**					**12**					
16	**P**	22	13	17			4		26					15		3			17
	G	**22**	**13**	**17**			**5**		**1**					**16**					**1**
17	**P**					4		17		4	6	17	17		34		7	4	
	G																		
18	**P**							14		6	3	12	22		15		8	18	
	G							**14**		**6**	**3**	**12**	**1**		**13**		**7**	**17**	

Table S1.2: Descriptive statistics for the six traits, where N is the number of observation having both phenotypes and genotypes, SD is the standard deviation, SE is the standard error and CV is the coefficient of variation expressed as percentage.

	N	Unit	Min	Max	Median	Mean (SE)	SD	CV%
BWH	1119	g	115.60	802.80	390.20	407.31 (3.84)	128.44	31.53
BL	1119	cm	14.10	28.00	22.40	22.38 (0.07)	2.25	10.05
BD	1119	cm	5.00	12.00	8.80	8.89 (0.03)	1.03	11.58
BT	1119	mm	12.90	59.70	40.50	40.70 (0.14)	4.55	11.17
FW	1119	g	20.10	342.60	136.60	143.83 (1.56)	52.32	36.38
FY	1119	%	15.24	50.53	33.15	32.83 (0.09)	3.13	9.54

Table S1.3: Mean values of the genomic relationship matrices constructed with NOIA and HWE approaches

	HWE			NOIA		
	Overall	Diagonal	Off-diagonal	Overall	Diagonal	Off-diagonal
G	0.00079	0.8847333	-0.000791354	0.000893	1	-0.000894454
D	0.038489	0.9250984	0.03690302	0.000893	1	-0.000894454
k(G#G)	0.009216	0.7868253	0.007825131	0.011713	1	0.009945195
k(G#D)	0.002971	0.8211928	0.001507525	0.003311	1	0.001528032
k(D#D)	0.00513	0.8581108	0.003604039	0.004223	1	0.002441192

Table S1.4: Transformation of the variances on a similar scale based on the relationship matrices. The additive genetic variance (σ^2_A) and the heritability (h^2) were obtained from A model. No individual homozygosity was fitted in these models. The transformed variances and ratio are marked by * and was scaled by (Mean (diagonal) – Mean) of the the corresponding relationship matrices (Table S1.3) as per Legarra (2016)[1]. In "Ped", genomic relationship matrix was replaced by pedigree relationship matrix obtained using 3 generations of pedigree. The mean and mean(diagonal) of the relationship between 1119 individuals from pedigree relationship matrix were 0.2757587 and 1 respectively.

Traits	Assumption	σ^2_A	SE	σ^2_A*	SE	h^2	SE	h^{2*}	SE
BWH	**NOIA**	1242	300	1241	300	0.18	0.04	0.18	0.04
	HWE	1404	339	1241	300	0.20	0.04	0.18	0.04
	Ped	3489	1391	2556	1019	0.44	0.14	0.32	0.10
BD	**NOIA**	0.13	0.03	0.13	0.03	0.24	0.04	0.24	0.04
	HWE	0.14	0.03	0.13	0.03	0.27	0.05	0.24	0.04
	Ped	0.31	0.12	0.23	0.09	0.50	0.15	0.36	0.11
BL	**NOIA**	0.40	0.12	0.40	0.12	0.13	0.04	0.13	0.04
	HWE	0.46	0.14	0.40	0.12	0.14	0.04	0.12	0.03
	Ped	1.20	0.50	0.88	0.37	0.33	0.12	0.24	0.09
FW	**NOIA**	177	47	177	47	0.15	0.04	0.15	0.04
	HWE	200	53	177	47	0.17	0.04	0.15	0.04
	Ped	536	215	393	158	0.39	0.13	0.28	0.09
BT	**NOIA**	1.70	0.44	1.69	0.44	0.17	0.04	0.17	0.04
	HWE	1.92	0.50	1.69	0.44	0.19	0.04	0.17	0.04
	Ped	1.83	0.92	1.34	0.68	0.18	0.08	0.13	0.06
FY	**NOIA**	1.76	0.41	1.76	0.41	0.19	0.04	0.19	0.04
	HWE	1.99	0.46	1.76	0.41	0.21	0.04	0.19	0.04
	Ped	2.65	1.17	1.94	0.86	0.26	0.10	0.19	0.07

[1] Legarra A. Comparing estimates of genetic variance across different relationship models. Theor Popul Biol. Elsevier; 2016;107:26–30.

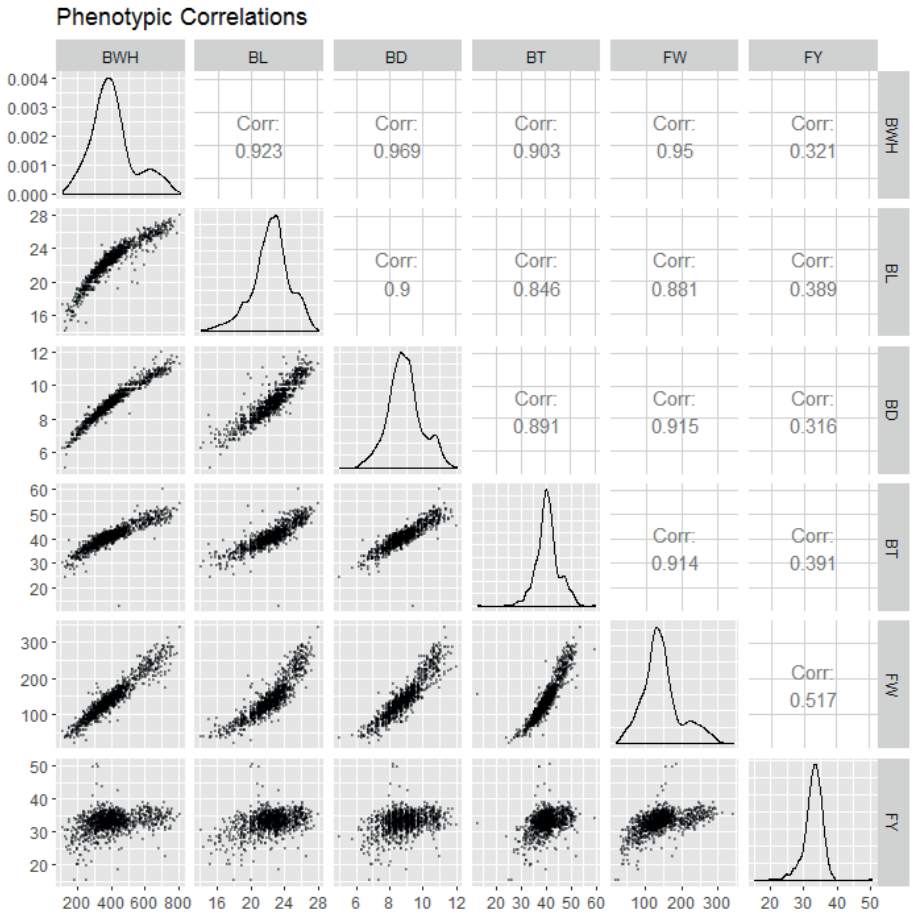


Figure S1.1: Scatterplots, histograms and the correlations of the 6 traits studied.

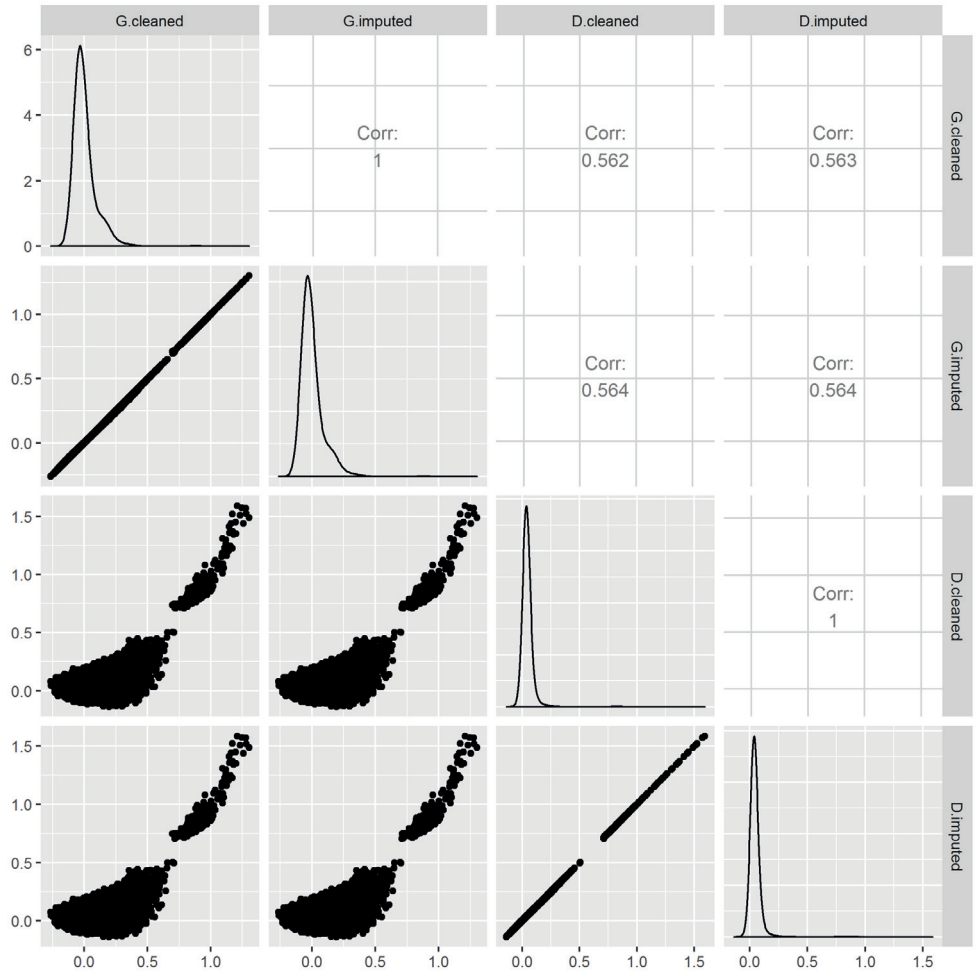


Figure S1.2: Scatterplot and correlation of the additive and dominance relationships using imputed genotypes (G.imputed and D.imputed respectively) and without imputed genotypes (i.e. with some missing genotypes, named as G.cleaned and D.cleaned respectively).

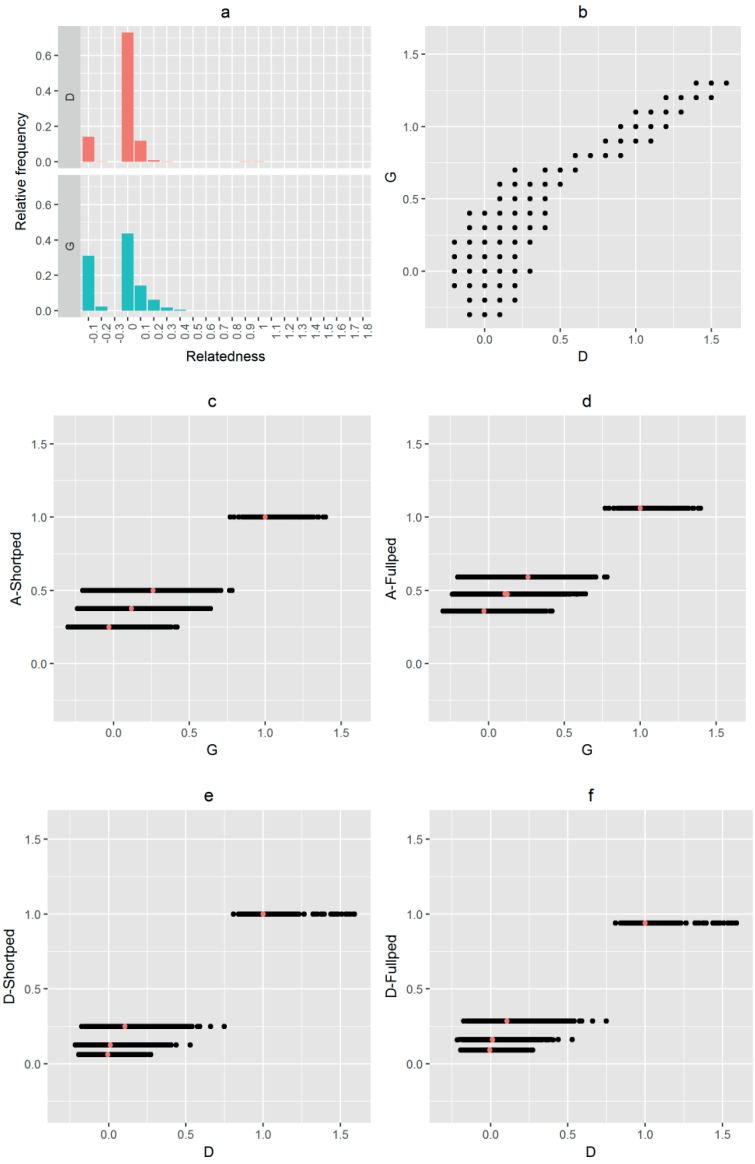


Figure S1.3: Scatterplots for different additive and dominance relationships. Mean of G and D are shown as orange dot in the respective plots **a)** Relative frequency plots for genomic relationships (G) and dominance relationships using genomic information (D). **b)** Scatterplot between G and D (correlation = 0.47). **c)** Scatterplot between G (using NOIA approach) and the pedigree relationships using 3 generations of pedigree (A-Shortped) (correlation = 0.72). **d)** Scatterplot between G and the pedigree relationships using 20 generations of pedigree (A-Fullped) (correlation = 0.72). **e)** Scatterplot between D and the dominance relationships using 3 generations of pedigree (D-Shortped) (correlation =0.65). **f)** Scatterplot between D and the dominance relationships using 20 generations of pedigree (D-Fullped) (correlation = 0.62).

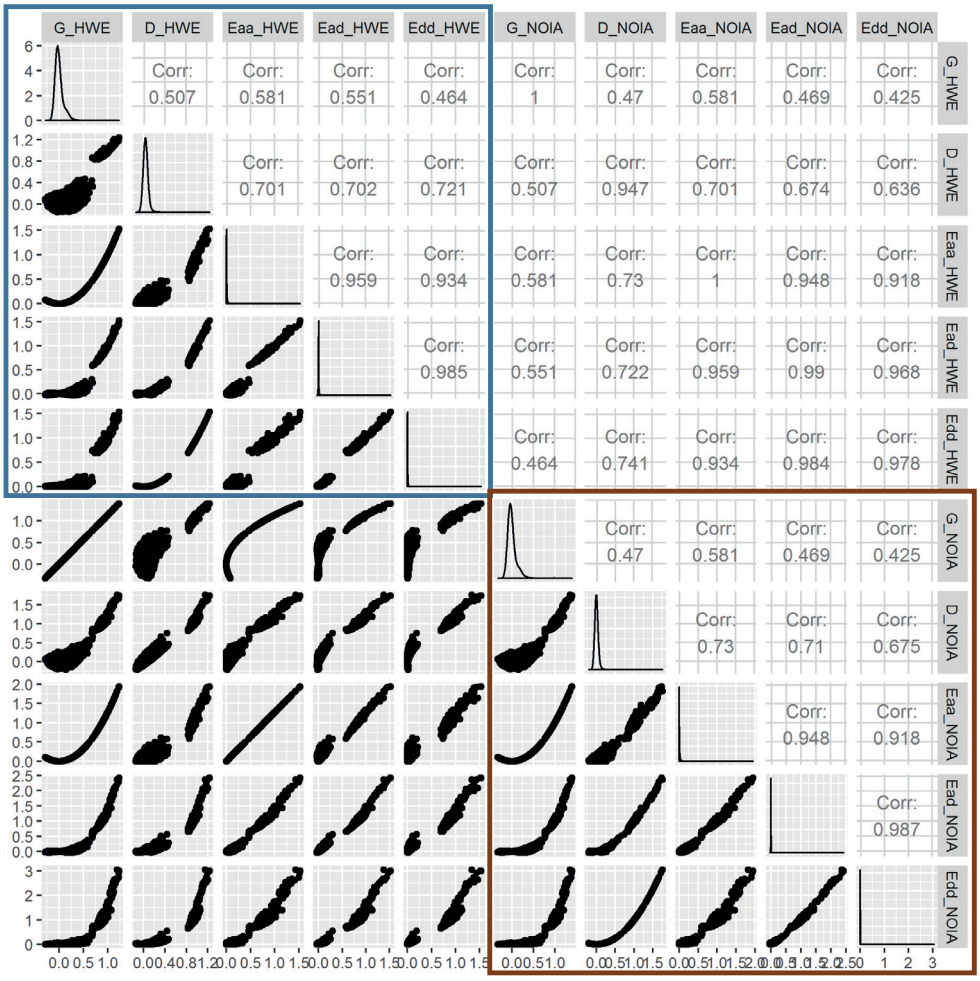


Figure S1.4: Scatterplots for different additive dominance and epistasis relationships using NOIA (inside brown box) and HWE assumption approaches (inside blue box).

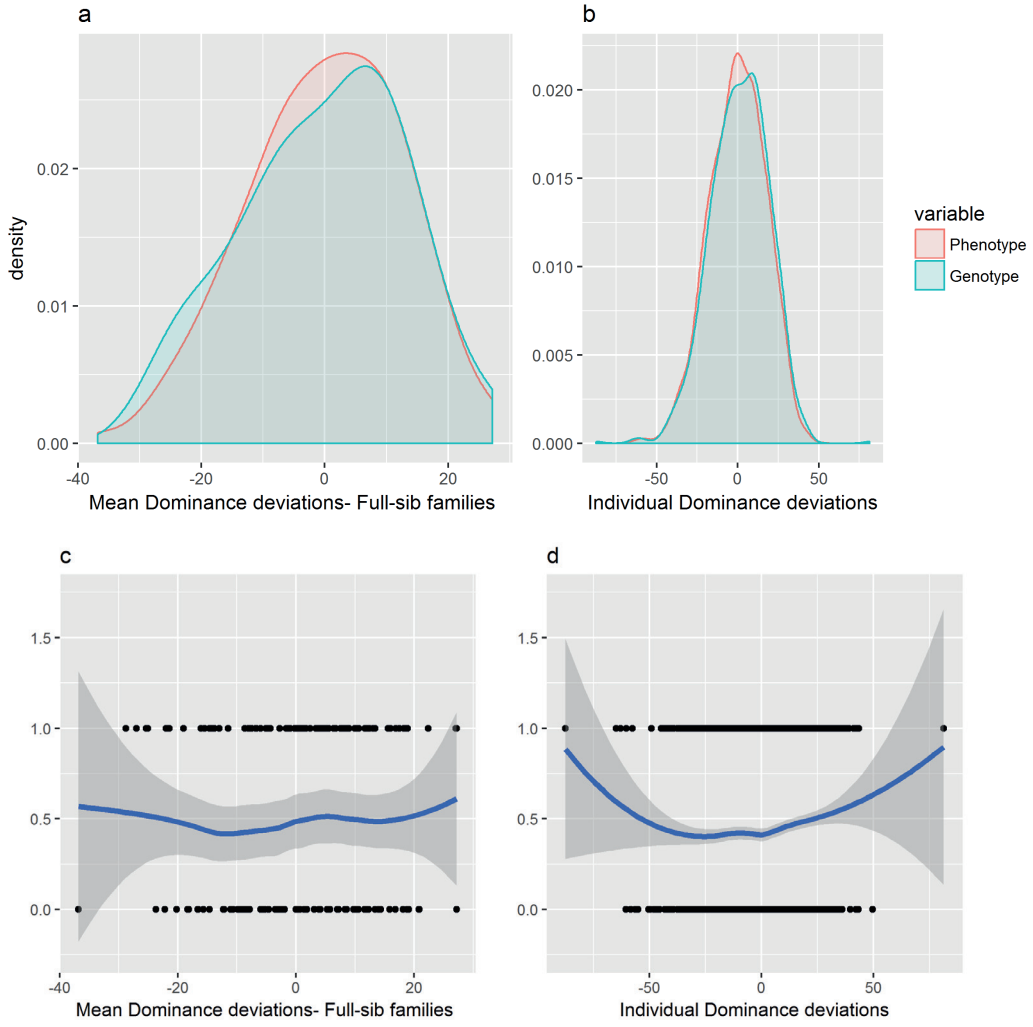


Figure S1.5: **a)** and **b)** Density plots showing the dominance deviations for BWH between the animals having phenotypes and those that were selected from them to be genotyped. **c)** and **d)** Scatterplot and LOESS regression between the selected and non-selected individuals. The selected individuals were coded as 1 and the non-selected individuals were coded as 0. The dominance deviations were obtained from the ADM model given in Joshi et al. (2018)[2].

[2] Joshi R, Woolliams J.A., Meuwissen T.H.E., Gjøen H.M.. Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity (Edinb)* [Internet]. Nature Publishing Group; 2018 Jan 16 [cited 2018 Jan 16];1. Available from: http://www.nature.com/articles/s41437-017-0046-x

Supplementary 3:

Assumptions on the nature of non-additive genetic variance and the impact on estimates of additive genetic variance

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Joshi et al. (2018)[1] showed how the variance components of the basic factorial mating design used in this study were translated into estimates of maternal, dominance and additive variances related to a pedigree. The basic factorial mating design had 3 core components (V_{Sire}, V_{Dam}, and V_{Fsib}) which can be related to the covariances (C) between individuals, *i* and *j*, assuming a mean is fitted to the population.

i, *j* no common parent (U), $C_U = 0$ (1)

i, *j* paternal half-sibs (PHS), $C_{PHS} = V_{Sire}$ (2)

i, *j* maternal half-sibs (MHS), $C_{MHS} = V_{Dam}$ (3)

i, *j* full-sibs (FS), $C_{FS} = V_{Sire} + V_{Dam} + V_{Fsib}$ (4)

For this population *i* and *j* were in generation 22, and Joshi et al. (2018)[1] published the main results with a base set at generation 20.

Dominance. Assuming the non-additive genetic variation was primarily arising from dominance then Joshi et al. (2018)[1] showed:

$$C_U = (4\sigma^2_A + \sigma^2_D)/16 \quad (5)$$

$$C_{PHS} = (6\sigma^2_A + 2\sigma^2_D)/16 \quad (6)$$

$$C_{MHS} = (6\sigma^2_A + 2\sigma^2_D)/16 + \sigma^2_M \quad (7)$$

$$C_{FS} = (8\sigma^2_A + 4\sigma^2_D)/16 + \sigma^2_M \quad (8)$$

The fitted mean will account for the genotypic drift from the base generation, which is represented by C_U, and Equation 5 can be subtracted from the (6), (7) and (8).

$$C_{PHS} = (2\sigma^2_A + \sigma^2_D)/16 \quad (9)$$

$$C_{MHS} = (2\sigma^2_A + \sigma^2_D)/16 + \sigma^2_M \quad (10)$$

$$C_{FS} = (4\sigma^2_A + 3\sigma^2_D)/16 + \sigma^2_M \quad (11)$$

[1] Joshi R, Woolliams J.A., Meuwissen T.H.E., Gjøen H.M.. Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity (Edinb)* [Internet]. Nature Publishing Group; 2018 Jan 16 [cited 2018 Jan 16];1. Available from: http://www.nature.com/articles/s41437-017-0046-x

Solving these equations and equating them to (2) to (4) results in the following:

- σ^2_M is estimated as $C_{MHS} - C_{PHS}$.
- σ^2_D is estimated as $16(C_{FS}–C_{PHS}–C_{MHS}) = 16V_{Fsib}$
- σ^2_A is estimated as $16C_{PHS}-8(C_{FS}-C_{MHS}) = 8(V_{Sire} - V_{Fsib})$.

Epistasis[2]. Consider now an assumption that the non-additive genetic variation was primarily arising from A#A, denoted σ^2_I :

$$C_U = (16\sigma^2_A + 4\sigma^2_I)/64 \quad (5)$$

$$C_{PHS} = (24\sigma^2_A + 9\sigma^2_I)/64 \quad (6)$$

$$C_{MHS} = (24\sigma^2_A + 9\sigma^2_I)/64 + \sigma^2_M \quad (7)$$

$$C_{FS} = (32\sigma^2_A + 16\sigma^2_I)/64 + \sigma^2_M \quad (8)$$

As with dominance the fitted mean removes C_U and this is subtracted from remaining covariances.

$$C_{PHS} = (8\sigma^2_A + 5\sigma^2_I)/64 \quad (6)$$

$$C_{MHS} = (8\sigma^2_A + 5\sigma^2_I)/64 + \sigma^2_M \quad (7)$$

$$C_{FS} = (16\sigma^2_A + 12\sigma^2_I)/64 + \sigma^2_M \quad (8)$$

The solutions to these equations are:

- σ^2_M is estimated as $C_{MHS} - C_{PHS}$.
- σ^2_I is estimated as $32(C_{FS}–C_{PHS}–C_{MHS}) = 32V_{Fsib}$
- σ^2_A is estimated as $28C_{PHS}-20(C_{FS}-C_{MHS}) = 8V_{Sire} - 20V_{Fsib} = 8(V_{Sire} - V_{Fsib}) – 12V_{Fsib}$.

Therefore the estimate of σ^2_A from this design is reduced when the non-additive variation is assumed to be additive-by-additive epistasis rather than dominance, and this reduction is of the order of $3/8\ \sigma^2_I$.

[2] Cockerham CC. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics.* Genetics Society of America; 1954;39(6):859.

Supplementary 4: Impact of inbreeding depression in the models

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Both the models with HWE and NOIA approaches were fitted without individual homozygosity as the covariate to account for the impact of the inbreeding depression in the models. The summary of the variance parameters are presented in the tables below.

Table S2.1: Heritabilities, ratio and phenotypic variance, for the models of best fit for different traits. The relationship matrices were constructed with HWE approach[1]. Models were not fitted with individual homozygosity as the covariate.

HWE approach - Without individual homozygosity											
Traits	**Model**	h^2	se	e_{aa}^2	se	H^2	se	m^2	se	phenvar	se
BD	AME	0.17	0.05	0.19	0.11	0.36	0.10	0.08	0.05	0.58	0.04
BWH	AME	0.11	0.04	0.22	0.11	0.33	0.10	0.08	0.05	7540	548
BL	AM	0.10	0.03					0.08	0.05	3.41	0.22
FW	AM	0.11	0.04					0.08	0.05	1252	82
BT	A	0.20	0.04							9.96	0.50
FY	A	0.21	0.04							9.45	0.47

Table S2.2: Heritabilities, ratio and phenotypic variance, for the models of best fit for different traits. The relationship matrices were constructed with NOIA approach[1]. Models were not fitted with individual homozygosity as the covariate.

NOIA approach - Without individual homozygosity											
Traits	**Model**	h^2	se	e_{aa}^2	se	H^2	se	m^2	se	phenvar	se
BD	AME	0.16	0.04	0.16	0.09	0.32	0.09	0.08	0.05	0.54	0.04
BWH	AME	0.10	0.04	0.19	0.10	0.29	0.09	0.09	0.05	7110	499
BL	AM	0.09	0.03					0.08	0.05	3.38	0.22
FW	AM	0.10	0.03					0.08	0.05	1236	80
BT	A	0.18	0.04							9.74	0.46
FY	A	0.19	0.04							9.23	0.44

[1] Vitezica ZG, Legarra A, Toro MA, Varona L. Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. Genetics [Internet]. 2017 Jul [cited 2018 Jan 29];206(3):1297–307. Available from: http://www.ncbi.nlm.nih.gov/pubmed/28522540

Table S2.3: Literature review for inbreeding depression in some species of aquaculture. The inbreeding depression is expressed as the percentage decrease in the trait value per 10% increase in the inbreeding coefficient.

Species	Trait	Inbreeding depression
Atlantic salmon[2]	BW	-0.6 to -2.6%
Rainbow trout[3]	BW	-2.3%
Rainbow trout[4]	BWH	-1.6 to -5.0%
Coho salmon[5]	BWH	-1.5% to -1.7%

BW- Body Weight, BWH- Body Weight at Harvest

[2] Rye M, Mao IL. Nonadditive genetic effects and inbreeding depression for body weight in Atlantic salmon (Salmo salar L.). Livest Prod Sci. Elsevier; 1998;57(1):15–22.

[3] Hu G, Wang C, Da Y. Genomic heritability estimation for the early life-history transition related to propensity to migrate in wild rainbow and steelhead trout populations. Ecol Evol [Internet]. 2014 Apr 20 [cited 2015 Oct 16];4(8):1381–8. Available from: http://doi.wiley.com/10.1002/ece3.1038

[4] Pante MJR, Gjerde B, McMillan I. Effect of inbreeding on body weight at harvest in rainbow trout, Oncorhynchus mykiss. Aquaculture [Internet]. Elsevier; 2001 Jan 15 [cited 2018 Sep 4];192(2–4):201–11. Available from: https://www.sciencedirect.com/science/article/pii/S0044848600004671

[5] Neira R, Díaz NF, Gall GAE, Gallardo JA, Lhorente JP, Manterola R. Genetic improvement in Coho salmon (Oncorhynchus kisutch). I: Selection response and inbreeding depression on harvest weight. Aquaculture [Internet]. Elsevier; 2006 Jun 30 [cited 2018 Sep 4];257(1–4):9–17. Available from: https://www.sciencedirect.com/science/article/pii/S0044848606001839

Supplementary 5: Impact on the genetic evaluation

Genomic dissection of maternal, additive and non-additive genetic effects for growth and carcass traits in Nile tilapia

R Joshi, JA Woolliams, THE Meuwissen and HM Gjøen

Both the models with HWE assumption and NOIA approaches were fitted without individual homozygosity as the covariate to account for the impact of the inbreeding depression in the models (marked as *). The comparison is made between A* model (fitting only additive genetic variation) and the higher order alternative models (fitting individual homozygosity, dominance, additive-by-additive epistasis and maternal environmental effect, single or in combination), therefore the 0 for the A* model is by definition. There was no much difference in the rankings between the HWE or NOIA approaches, so the results of NOIA approach is only shown.

Table S5.1: The impact of model choice for the top 100 animals after ranking was based on EBVs. The boxes are color coded from dark grey to white so that the darker color signifies higher values.

Traits	**Models without individual homozygosity**							
	A*	**AD***	**ADE***	**ADM***	**ADME***	**AE***	**AM***	**AME***
BD	0	2	8	30	31	8	30	31
BWH	0	0	6	40	39	6	40	39
BL	0	1	10	54	52	10	54	52
FW	0	0	7	43	44	7	43	44
BT	0	1	5	23	19	5	22	19
FY	0	0	4	15	14	4	15	14
	Models with individual homozygosity							
	A	**AD**	**ADE**	**ADM**	**ADME**	**AE**	**AM**	**AME**
BD	9	10	8	29	31	8	29	31
BWH	8	8	10	40	37	10	40	37
BL	12	12	14	53	52	14	53	52
FW	5	5	6	40	41	6	40	41
BT	2	3	4	23	18	4	23	18
FY	4	4	4	14	15	4	14	15

Adding only individual homozygosity as the measure of inbreeding depression made minor differences in the ranking of the 100 best animals, with the change of 4-12% of the animals in the top 100 list across various traits. Similarly, adding only epistatic term also made minor difference in the top 100 list based on EBVs, with the change of 4-10% of

the animals across various traits. The impact of the epistatic effects on the ranking based on EBVs was not consistent with the significance of this effect for the trait. For example, though the best-fit model for the trait BL did not have epistatic effect, including this effect in the model for this trait changed the maximum number of animals in the list compared to other traits where this effect was significant.

Whereas, including maternal term in the models was found to effect the ranking based on EBVs, with more than 50% of the animals being different in the list. For those traits where the best-fit models included maternal variances, the impact of the maternal effect was bigger, than those traits where the maternal effect was not significant. These results are similar to that using pedigree[1].

[1] Joshi R, Woolliams J.A., Meuwissen T.H.E., Gjøen H.M.. Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. Heredity (Edinb) [Internet]. Nature Publishing Group; 2018 Jan 16 [cited 2018 Jan 16];1. Available from: http://www.nature.com/articles/s41437-017-0046-x

4. General Discussion

Prof. Harald Skjervold's at NMBU (former NLH) first involvement with Atlantic salmon was to use it as a model species to disentangle additive and non-additive genetic effects in a factorial design similar to the one we have applied in the present project (Gjedrem, 2005). The external fertilization, which easily could be applied in fish, was unattainable with other production animals. This shows that non-additive genetic effects were thought to be an important source of genetic variation in aquaculture since the very beginning. One of the main conclusion and the future research priority of the Aquaculture Genomics, Genetics and Breeding Workshop held in Auburn, Alabama, USA in 2016 (Abdelrahman *et al.*, 2017) was to try to utilize heterosis in different aquaculture species, which shows that this is still relevant. Hence, this project was also to some extent motivated by the huge difference in profit between breeding programs that exploit both non-additive and additive vs. those that exploit only additive genetic effects (Bichard, 1977; Weller, 1994). For example, pigs and chickens rely on the additivity, but they exploit non-additivity as well (Bell *et al.*, 1952). The thesis presents the papers in the order of the progression of work: first we use classical methods utilizing pedigree to partition the variances and quantify the non-additive genetic effects, then we develop the genomic resources (SNP-array and linkage maps) to allow deeper analyses, and finally utilize these genomic resources to partition the variances and quantify the non-additive genetic effects. Many relevant issues have already been addressed in the discussion sections of the respective papers and will not be repeated here. Thus, this general discussion will focus on following main points, with the aim to provide broader general perspective and possible future areas of research than what has already been given in the papers.

4.1. Experimental Design

Most conventional tilapia aquacultural breeding designs cause confounding of various effects; making it difficult to separate various source of variances during the genetic evaluation. In this experiment, the source population was part of a breeding nucleus and various modifications of the regular commercial Nile tilapia practice were done to separate the additive, maternal and non-additive genetic effects accurately, as shown in Table 2 (See methodology in paper I for details).

Table 2: Modifications in the regular commercial Nile tilapia aquaculture practice

Regular commercial practice	**Confounding**	**Modifications**
Hierarchical mating /nested design	Difficult to separate maternal and non-additive genetic effects	Factorial design with reciprocal cross
Natural mating and mouth brooding	Maternal environmental effect and full-sib family effect	Artificial breeding (stripping the eggs and milt directly from the genital papilla and fertilizing them artificially in mixing containers)
Full-sib families are kept separate until they are pit-tagged (for about 5-7 weeks)	Tank effect, maternal environmental effect and full-sib family effect	All the eggs were kept together and were reared together in the same environment. Parentage assignment was done with the aid of microsatellites

4.2. Models used

The prediction of breeding values is the important core of any breeding program, and we depend on the models used to get these values. As with the "Garbage in- garbage out" paradigm, the inappropriate data with perfect model or the perfect data with inappropriate model leads to wrong conclusion. Also, it would be impossible to get the perfect data and perfect model in the practical world. Thus, practically we try to have an experimental design (as above) to generate the best data for an experiment and utilize the best available models (with their own assumptions).

In animal breeding and genetics, BLUP (Best Linear Unbiased Prediction) models (Henderson, 1949) have been widely utilized, which has evolved from simpler sire models to the animal models and on to the genomic models. BLUP assumes that the appropriate variance components are known (Mrode, 2014). Paper I and III utilizes

different mixed linear models using residual maximum likelihood (REML) to partition the additive and non-additive genetic variances.

Paper I compares the sire variances and full-sib variances obtained from simple diallel model, with the standard animal models with dominance relationship matrices (coefficient of fraternity) (Harris, 1964; Smith and Mäki-Tanila, 1990; Lynch and Walsh, 1998). Use of fraternity matrices for dominance (computed from pedigree information) is an approximation of the full dominance model (Shaw and Woolliams, 1999). The presence of inbreeding together with dominance increases the deviation from the HWE and generates covariances between additive and dominance effects (de Boer and Hoeschele, 1993). This can be written as the computationally simpler equivalent model (Fernandez *et al.*, 2017), which ignores two terms related to the inbreeding depression:

$$Var\begin{pmatrix}\boldsymbol{a}\\ \boldsymbol{d}\end{pmatrix} = \begin{pmatrix}\boldsymbol{A}\sigma_A^2 & \boldsymbol{C}\sigma_{A,D_I}\\ \boldsymbol{C}'\sigma_{A,D_I} & \boldsymbol{D_B}\sigma_{D_B}^2 + \boldsymbol{D_I}\sigma_{D_I}^2\end{pmatrix}$$

where, **a** is the vector of random additive genetic effects; **d** is the vector of random dominance effects; σ^2_A and σ^2_{DB} are the additive and dominance genetic variances in the base population (non-inbred population mated at random); σ^2_{DI} is the dominance genetic variance in a fully inbred population (with the same allelic frequencies as the base; $\sigma_{A,DI}$ is the covariance between the respective additive and dominance effects; **A** is the numerator relationship matrix generated from the pedigree (twice the probability of IBD of two alleles across two individuals); $\mathbf{D_I}$ is inbred dominance relationship matrix (probability of both identical genotypes across individuals and identical alleles within individuals); $\mathbf{D_B}$ is the non-inbred dominance relationship matrix (probability of identical genotypes across individuals but distinct alleles within individuals); and **C** is the additive-dominance relationship matrix (probability of three alleles identical across the two pairs of two individuals.

Hence, the fraternity relationship matrices used in paper I excludes term that arises in the presence of inbreeding with dominance, i.e. the terms $\boldsymbol{D_I}\sigma^2_{D_I}$ and the co-variances in the above variance structure were assumed to be zero. But the relatively low value of the inbreeding coefficient in our population (compared to the fully-inbred

individuals where these full models have been utilized), may not cause too serious problems and biasness.

These BLUP parametric methods assume that the dominance effects have the mean of zero and the symmetric distribution of the variance (Falconer *et al.*, 1996; Mrode, 2014). Whereas, under the directional dominance these assumptions do not hold true. One example to overcome this shortcoming is to use the regression of homozygosity as the covariate in the model, as we did in paper III.

Different genomic regions in the genome has been shown to contribute differently to the inbreeding depression and overall heterosis (for example: (Pryce *et al.*, 2014)). Hence, it is difficult to model these properties using parametric methods like linear mixed models and non-parametric models could be an option (For example Bayes D, as in (Bennewitz *et al.*, 2017)), though the implementation is complex. The GBLUP approaches used here suggest the linear relationship between the genotypes and phenotypes, which might not be true when we want to model epistasis (Gianola *et al.*, 2006). Hence, appropriate non-linear models with easy implementation needs to be developed.

4.3. New genomic resources for Tilapia

The first Aquaculture Genomics Workshop, Dartmouth, Massachusetts, United States of America in 1997 is considered a milestone for starting aquaculture genomics, and tilapia was one of the six species selected for genomic research (Liu, 2017). At present, we have a spectrum of genomic resources available for different aquaculture species, ranging from restriction-site association DNA (RAD)/ double digested RAD (ddRAD) and only a little knowledge of genome structure or gene content (for example Yellow Croaker), to high-density arrays with significant genome annotation (Atlantic salmon - *Salmo salar*). High-resolution data (such as we can get from *S. salar*) has more chance to reveal biology underlying genotype-phenotype links and greater chance to help us select more efficient for complex traits.

The Aquaculture Genomics, Genetics and Breeding Workshop held in Auburn, Alabama, USA in 2016 has reviewed and compared the status of aquaculture genomics for major aquaculture species (Abdelrahman *et al.*, 2017). Despite having two genome assemblies (see (NCBI, 2018)) and four linkage maps of varying resolutions (Kocher

et al., 1998; Lee *et al.*, 2005; Guyon *et al.*, 2012; Palaiokostas *et al.*, 2013), the workshop highlighted the lack of utilization of available genome tools and technologies in tilapia breeding programs (See Figure 1 at page 15 of (Abdelrahman *et al.*, 2017)). To bridge this gap, Paper II deals with the development of two important genomic resources, 58K SNP-array (Onil50 array) and the high-density linkage map, for Nile tilapia.

Priority of utilizing SNP array and the linkage maps in aquaculture has been in utilizing these resources on selection for disease and parasitic resistance. Hence, one of the major utilities of Onil50 array will be genome wide association studies (GWAS), making it preferable to have the SNPs at equivalent distances across the genome. Hence, even though a very thorough and stringent approach was used to develop the array, in Paper II it has been recommended to increase the coverage and density of the Onil50 array, so that it becomes comparable to the recently available SNP-arrays in another important aquacultural species, the 930K in Salmon (Lien *et al.*, 2016).

Besides increasing the density of the SNP arrays, another future perspective in tilapia breeding might be to utilize the Whole Genome Sequence (WGS) data, i.e. to utilise all base-pairs on the genome that show variation in the population. Assuming the sequencing of enough animals, it is believed that the data will contain nearly all causal mutations contributing to the genetic variance. This will have a direct impact on short term breeding by increasing the power to detect and localise causal variants (Meuwissen and Goddard, 2010; MacLeod *et al.*, 2013; Druet *et al.*, 2014; van Binsbergen, 2017). Whereas the long term breeding goals might be to use the WGS information combined with the newer methods like gene editing and RNA Seq data for genomic prediction (Jenko *et al.*, 2015).

The Build 1 linkage map (see methodology in Paper II) was used to anchor, order and orient the scaffolds in the available O_niloticus_UMD1 genome assembly. Although this is not the part of this thesis, it shows one of the utilization of the linkage map we produced. In short, this linkage map helped to identify 22 mis-assemblies, anchored a total of 906.6 Mbp of the O_niloticus_UMD1 assembly and was used to produce a new version of the assembly: O_niloticus_UMD_NMBU assembly (Conte *et al.*, 2018). An important contribution was that the linkage map helped to join the LG3a and LG3b into a single linkage group, LG3, in the assembly.

Availability of genome sequence and the establishment of gene-editing methods (Li *et al.*, 2013, 2014; Brawand *et al.*, 2014) make Nile tilapia an ideal model species for the experiments related to the gene-editing. It is expected that the future research will focus on creating different disease resistance, salt tolerant and temperature tolerant lines of Tilapia, and we believe that the development of SNP-array and linkage map is a enabling the first step towards this goal.

4.4. Maternal effect

Maternal effects can either be genetic or non-genetic sources of variation among the offspring acquired from the mother. For example: the size of the egg is an environmental factor for the offspring, whereas for the mother it can be genetic so that it gets transferred to the offspring. These environmental sources of variation in offspring has been shown to origin from either of the parents (Immler, 2018), but female influences have been considered more significant, in Nile tilapia as in many other fish species, as the female provides the eggs with nutrients, hormones, cytoplasm, and mitochondrial DNA. Additionally, in Nile tilapia the female provides the brooding environment and chooses the place to deposit them. Hence, mother is a more likely source of the variance, at least during the initial stages of development (Heath *et al.*, 1999).

In the beginning, maternal effects were viewed as an experimental noise, but eventually this has been converted as an opportunity to breed for separate maternal lines (Smith, 1964). In aquaculture (i), the size of the dam is correlated to the number of eggs produced (i.e. reproductive potential of the dam) (Marshall *et al.*, 1998); (ii) variation in the age and size of the dam is correlated to the qualitative changes in eggs (Solemdal, 1997); (iii) size of egg is correlated with the size of hatchling (Chambers *et al.*, 1989); (iv) negative correlation between egg size and number (Smith and Fretwell, 1974); (v) female mouth brooding is an important natural phenomena responsible for maternal environmental effect in Nile tilapia, and it has been found that the low offspring density in the pouch favours growth (Watanabe and Watanabe, 2002); (vi) Transfer of maternal immunity to eggs and embryos through yolk (Mor and Avtalion, 1990; Swain and Nayak, 2009). The detailed review on the causes, modes and mechanisms of maternal effects in fishes can be found in Green (2008).

The influence of these maternal effects on the traits of offspring generally declines with the gradual development of the offspring, because with time and life history, the influence on offspring's trait is gradually taken over by its genetic architecture (Atchley and Zhu, 1997; Heath and Blouw, 1998) and the maternal effects in fishes are postulated to be negligible after the early juvenile stages (Heath and Blouw, 1998). But in our experiment, we found that the maternal effects were still responsible for four out of six commercial growth and carcass traits (Paper I and III), and was found to explain around 10% of the phenotypic variance ($P<0.05$). Since these maternal ratios are reported for the first time in Nile tilapia, a comparison from other studies was not available within the species.

This result challenges the hypothesis of negligible maternal effects after the juvenile stages. One possible explanation is that this maternal effect might be the manifestation of mitochondrial effects (Huizinga *et al.*, 1986; Rothschild and Ollivier, 1987), assuming negligible maternal effects related to size and quality of the eggs at the harvest age of Nile tilapia. Another possibility is that maternal effects might be species specific (Benowitz *et al.*, 2015) and all the maternal effects are still present at harvest age. The origin of maternal effects are the substantially larger variances due to dams rather than sires (Falconer *et al.*, 1996). It has also been shown that incorrect and incomplete pedigrees cause biasness in the calculation of the maternal ratio (Roughsedge *et al.*, 2001). But before concluding anything, future research should be directed on manifestation of maternal effects in the different ontogenetic stages of Nile tilapia and some other fish species. Availability of the genomic resources with the SNPs covering the mitochondrial genome (Paper II) makes it very interesting to search for the genes located in the mitochondria (Gonçalves *et al.*, 2014) responsible for the manifestations of these commercial traits in Nile tilapia (and even study the gene expressions during different ontogenetic stages).

Including maternal component in the models was found to cause substantial impact in the genetic evaluation (Paper I and Supplementary 5 of Paper III). The next step would be to know if this is real or not in a normal population, and if it is truly maternal and to what extent the maternal effect is genetic. This will provide an opportunity for the creation of maternal lines in Tilapia breeding schemes.

4.5. Non-additive genetic effect

Most of the genetic evaluations ignore the non-additive genetic effects and are based on the hypothesis that the genes determining the phenotype act solely by adding their effects. This quantification of the role of genetics for any trait is done by calculating heritability, which is the ratio of genetic variance over phenotypic variance. It is hard to say that this method of calculation precisely gives the correct estimation of additive variance, as it has been well shown that the additive variance with this method contains some fragment of dominance and interactions (Falconer *et al.*, 1996) Therefore, it is unknown how much of the total genetic variance is precisely due to the additive component in Nile tilapia, though there have been efforts to separate the contribution of each variance term using modern technologies and advanced models in other species (Vinkhuyzen *et al.*, 2013).

The additional full-sib family variance (V_{Fsib}) over the sire and dam components in simple sire and dam model in the diallel design (Paper 1) under the dominance model can be interpreted (Falconer *et al.*, 1996) as;

$$V_{Fsib} = ¼\ V_D$$

where, V_D is the dominance variance.

In the presence of the dominance and epistasis, this is given by

$$V_{Fsib} = ¼V_D + ⅛V_{AA} + ⅛V_{AD} + ⅛V_{DD},$$

where, V_{AA}, V_{AD} and V_{DD} are additive-by-additive, additive-by-dominance and dominance-by-dominance epistatic variances. Hence, we have assumed in Paper I that the variances within the full-sib families are due to the dominance.

Due to the development of the genomic resources for Nile tilapia (paper II), we were able to use genomic models to partition the additive and non-additive genetic variances. It is much easier to calculate dominance and epistasis using genomics than pedigree, because in genomics we don't have to use the probabilities of identical genotypes, rather we use allele counts and the observations of heterozygote states of the SNPs. We fitted additive, dominance and epistatic models using genomics with individual homozygosity as the linear covariate and found that the V_{Fsib} was almost

entirely additive-by-additive epistatic variance, unlike dominance as concluded in paper I, with the presence of inbreeding depression. This might be due to the infinitesimal property of the dominance effects. It has been shown that under an infinitesimal model with both additive and dominance effects, with the increase in the number of loci, either the dominance variance tends towards zero or the inbreeding depression tends towards infinity (Falconer *et al.*, 1996; Toro and Mäki-Tanila, 2018). Further, we showed in Paper III that assuming the non-additive genetic effects to be epistasis, rather than dominance, affects the estimation of the additive genetic variance.

The GBLUP models use the genomic relationship matrices (GRMs) and there are different ways to construct these GRMs (VanRaden, 2008; Yang *et al.*, 2010; Speed and Balding, 2015; Vitezica *et al.*, 2017) directly affecting the partition of the variances. In Paper III we have shown that relaxing the assumption of HWE in the construction of the GRMs (as in (Vitezica *et al.*, 2017)) deflate the additive-by-additive epistatic ratio by approximately 20 % and 18%, and heritability by 6% and 10% for two traits body depth (BD) and body weight at harvest (BWH) respectively, when the estimates are not transformed to the base generation (Legarra, 2016). Further, the models we have used are not fully orthogonal in the presence of LD. So, development of models to account for LD in the population will help to partition the variances orthogonally and to get more accurate values (Hill and Mäki-Tanila, 2015).

Though the magnitude of the non-additive genetic effects were found to be large (but also with large standard errors) with differences between the full-sib families, it was concluded that it was not beneficial to go for reciprocal recurrent selection (RRS) to exploit the non-additive variation in this population, as additive-by-additive epistatic variance is believed to be converted to the additive genetic variance due to random drift and selection in the long run (Hill, 2017). Though the conclusion not to use RRS cannot be generalized for other effects, as this might not be valid if the future analysis shows that the maternal variance is genetic (possibility to utilize maternal lines), but this will still depend on the cost structures. Further, this also supports the claim that additive genetic variance is the major source of genetic variance for quantitative traits (Falconer *et al.*, 1996; Mackay, 2014; Hill, 2017), with an underlying epistatic genetic architecture. The distinction between the additive and epistatic variance is not that important if the aim

is to predict the short-term responses including estimation of heritability, prediction of breeding values or the phenotypes using the additive variances. However, Mackay (2014) has listed following points where it would be beneficial to know whether the additive variance is the function of underlying epistasis or not for a quantitative trait: (i) for functional dissection of the phenotype-genotype map (ii) determination of the genetic interaction networks (iii) prediction of long term responses to selection (both natural and artificial) (iv) prediction of the consequences of inbreeding and genetic drift. Hence, it would be helpful if the future research for non-additive genetic effects in this population would be directed to find the major genes or the regions responsible for the non-additive effects (Ravi *et al.*, 2011; Powell *et al.*, 2013; Lopes *et al.*, 2014; Bolormaa *et al.*, 2015; Jiang *et al.*, 2017), including large epistatic effects and to further utilize them in the breeding programs (Carlborg *et al.*, 2006; Große-Brinkhaus *et al.*, 2010). Compared to finding out the causal variants showing additive effect, it is difficult to find out the casual variants showing epistatic effects, as this will require huge number of observation, additional SNP coverage of the genome and huge computational power (Misztal, 1997).

5. General conclusion and recommendations

- Significant non-additive genetic effects were found for BWH and BD, which was entirely additive-by-additive epistatic effects in the presence of inbreeding depression. Rather than depending on the interactions of unknown genes contributing to the non-additive genetic variance, exploitation of this effect in the future in Nile tilapia breeding program might depend on finding out the genes or genomic regions associated with the heterosis for the traits.
- Significant maternal effects were found for four out of six traits (except BT and FY). The creation of maternal lines in Tilapia breeding schemes (with subsequent cross-breeding designs) may be a possibility if this variation is found to be heritable.
- Non-additive genetic effects were not found to cause substantial consequences on the rankings of the selection candidates. However, minor impacts were observed in the heritability and genetic evaluations. Unlike the non-additive genetic effects, including maternal component in the models was found to cause substantial consequences on the genetic evaluations.
- The development of SNP-Array and linkage map has opened a new door of the genomic era in Nile Tilapia. These resources have the potential to improve the genetic gain through genomic selection and to surpass the difficulties of selection of the invasive traits.

6. References

Abdelrahman H, ElHady M, Alcivar-Warren A, Allen S, Al-Tobasei R, Bao L, *et al.* (2017). Aquaculture genomics, genetics and breeding in the United States: current status, challenges, and priorities for future research. *BMC Genomics* **18**: 191.

Atchley WR, Zhu J (1997). Developmental quantitative genetics, conditional epigenetic variability and growth in mice. *Genetics* **147**: 765–776.

Bell AE, Moore CH, Bohren BB, Warren DC (1952). Systems of breeding designed to utilize heterosis in the domestic fowl. *Poult Sci* **31**: 11–22.

Bennewitz J, Edel C, Fries R, Meuwissen THE, Wellmann R (2017). Application of a Bayesian dominance model improves power in quantitative trait genome-wide association analysis. *Genet Sel Evol* **49**: 7.

Benowitz KM, Moody KJ, Moore AJ (2015). Are species differences in maternal effects arising from maternal care adaptive? *J Evol Biol* **28**: 503–509.

Bentsen HB, Eknath AE, Palada-de Vera MS, Danting JC, Bolivar HL, Reyes RA, *et al.* (1998). Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of Oreochromis niloticus. *Aquaculture* **160**: 145–173.

Bichard M (1977). Economic efficiency of pig breeding schemes: A breeding company view. *Livest Prod Sci* **4**: 245–254.

van Binsbergen R (2017). Prospects of whole-genome sequence data in animal and plant breeding.

de Boer IJM, Hoeschele I (1993). Genetic evaluation methods for populations with dominance and inbreeding. *Theor Appl Genet* **86**: 245–258.

Bolormaa S, Pryce JE, Zhang Y, Reverter A, Barendse W, Hayes BJ, *et al.* (2015). Non-additive genetic variation in growth, carcass and fertility traits of beef cattle. *Genet Sel Evol* **47**: 26.

Bouvet J-M, Makouanzi G, Cros D, Vigneron P (2015). Modeling additive and non-additive effects in a hybrid population using genome-wide genotyping: prediction accuracy implications. *Heredity (Edinb)* **116**: 146–157.

Brawand D, Wagner CE, Li YI, Malinsky M, Keller I, Fan S, *et al.* (2014). The genomic substrate for adaptive radiation in African cichlid fish. *Nature* **513**: 375.

Bulfield G, Siller WG, Wight PA, Moore KJ (1984). X chromosome-linked muscular dystrophy (mdx) in the mouse. *Proc Natl Acad Sci* **81**: 1189–1192.

Carlborg Ö, Jacobsson L, Åhgren P, Siegel P, Andersson L (2006). Epistasis and the release of genetic variation during long-term selection. *Nat Genet* **38**: 418.

Carlborg Ö, Kerje S, Schütz K, Jacobsson L, Jensen P, Andersson L (2003). A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Res* **13**: 413–421.

Chambers RC, Leggett WC, Brown JA (1989). Egg size, female effects, and the correlations between early life history traits of capelin, Mallotus villosus: an appraisal at the individual level. *Fish Bull* **87**: 515–523.

Cockerham CC (1954). An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. *Genetics* **39**: 859.

Comstock RE, Robinson HF, Harvey PH (1949). Breeding procedure designed to make maximum use of both general and specific combining ability. *Agron J*.

Conte MA, Joshi R, Moore EC, Nandamuri SP, Gammerdinger WJ, Clark FE, *et al.* (2018). Chromosome-scale assemblies reveal the structural evolution of African cichlid genomes. *bioRxiv*.

Darwin C (1859). *The Origin of Species by Means of Natural Election, Or the Preservation of Favored Races in the Struggle for Life*. AL Burt.

Dekkers JCM (2007). Marker-assisted selection for commercial crossbred performance. *J Anim Sci* **85**: 2104–2114.

DeStefano AL, Hoeschele I (1992). Utilization of dominance variance through mate allocation strategies. *J Dairy Sci* **75**: 1680–1690.

Dickerson GE (1969). Experimental approaches to utilizing breed resources. In: *Animal Breeding Abstract*, Vol 37, pp 191–202.

Druet T, Macleod IM, Hayes BJ (2014). Toward genomic prediction from whole-genome sequence data: impact of sequencing design on genotype imputation and accuracy of predictions. *Heredity (Edinb)* **112**: 39.

Dunn OJ, Clark V (1971). Comparison of tests of the equality of dependent correlation coefficients. *J Am Stat Assoc* **66**: 904–908.

Eknath AE, Tayamen MM, Palada-de Vera MS, Danting JC, Reyes RA, Dionisio EE, *et al.* (1993). Genetic improvement of farmed tilapias: the growth performance of eight strains of Oreochromis niloticus tested in different farm environments. *Aquaculture* **111**: 171–188.

Ertl J, Legarra A, Vitezica ZG, Varona L, Edel C, Emmerling R, *et al.* (2014). Genomic analysis of dominance effects on milk production and conformation traits in Fleckvieh cattle. *Genet Sel Evol* **46**: 40.

Esfandyari H, Bijma P, Henryon M, Christensen OF, Sørensen AC (2016). Genomic prediction of crossbred performance based on purebred Landrace and Yorkshire data using a dominance model. *Genet Sel Evol* **48**: 40.

Esfandyari H, Sørensen AC, Bijma P (2015). Maximizing crossbred performance through purebred genomic selection. *Genet Sel Evol* **47**: 16.

Falconer DS, Mackay TF (1996). 14 Inbreeding and Cross breeding. In: *Introduction to Quantitative Genetics*, Pearson Prentice Hall, pp 253–262.

Falconer DS, Mackay TF, Frankham R (1996). Introduction to Quantitative Genetics (4th edn). *Trends Genet* **12**: 280.

Fernandez EN, Legarra A, Martinez R, Sánchez JP, Baselga M (2017). Pedigree-based estimation of covariance between dominance deviations and additive genetic effects in closed rabbit lines considering inbreeding and using a computationally simpler equivalent model. *J Anim Breed Genet* **134**: 184–195.

Fisher RA (1919). XV.—The correlation between relatives on the supposition of Mendelian inheritance. *Earth Environ Sci Trans R Soc Edinburgh* **52**: 399–433.

Gallardo JA, Lhorente JP, Neira R (2010). The consequences of including non-additive effects on the genetic evaluation of harvest body weight in Coho salmon (Oncorhynchus kisutch). *Genet Sel Evol* **42**: 19.

Gianola D, Fernando RL, Stella A (2006). Genomic assisted prediction of genetic value with semi-parametric procedures. *Genetics*.

Gjedrem T (2005). Forskning, utvikling og engasjement innan akvakultur. In: Fra Veterinærlære til biovitenskap. : 236.

Gonçalves VF, Zai CC, Tiwari AK, Brandl EJ, Derkach A, Meltzer HY, *et al.* (2014). A hypothesis-driven association study of 28 nuclear-encoded mitochondrial genes with antipsychotic-induced weight gain in schizophrenia. *Neuropsychopharmacology* **39**: 1347.

Green BS (2008). Maternal effects in fish populations. *Adv Mar Biol* **54**: 1–105.

Große-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, Jüngst H, *et al.* (2010). Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc× Pietrain population. *Genet Sel Evol* **42**: 39.

Gupta M V, Acosta BO (2004). A review of global tilapia farming practices. *Aquac Asia* **9**: 7–12.

Guyon R, Rakotomanga M, Azzouzi N, Coutanceau JP, Bonillo C, D'Cotta H, *et al.* (2012). A high-resolution map of the Nile tilapia genome: a resource for studying cichlids and other percomorphs. *BMC Genomics* **13**: 222.

Habier D, Fernando RL, Dekkers JCM (2007). The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**: 2389–2397.

Hallauer AR, Eberhart SA (1970). Reciprocal full-sib selection. *Crop Sci* **10**: 315–316.

Harris DL (1964). Genotypic covariances between inbred relatives. *Genetics* **50**: 1319.

Heath DD, Blouw DM (1998). Are maternal effects in fish adaptive or merely physiological side effects. *Matern Eff as Adapt*: 178–201.

Heath DD, Fox CW, Heath JW (1999). Maternal effects on offspring size: variation through early development of chinook salmon. *Evolution (N Y)* **53**: 1605–1611.

Henderson CR (1949). Estimation of changes in herd environment. *J Dairy Sci* **32**: 706715.

Hill WG (2017). 'Conversion' of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response. *J Anim Breed Genet* **134**: 196–201.

Hill WG, Mäki-Tanila A (2015). Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *J Anim Breed Genet* **132**: 176–186.

Hill WG, Mulder HA (2010). Genetic analysis of environmental variation. *Genet Res (Camb)* **92**: 381–395.

Huizinga HA, Korver S, McDaniel BT, Politiek RD (1986). Maternal effects due to cytoplasmic inheritance in dairy cattle. Influence on milk production and reproduction traits. *Livest Prod Sci* **15**: 11–26.

Immler S (2018). The sperm factor: paternal impact beyond genes. *Heredity (Edinb)*.

Jenko J, Gorjanc G, Cleveland MA, Varshney RK, Whitelaw CBA, Woolliams JA, *et al.* (2015). Potential of promotion of alleles by genome editing to improve quantitative traits in livestock breeding programs. *Genet Sel Evol* **47**: 55.

Jiang J, Shen B, O'Connell JR, VanRaden PM, Cole JB, Ma L (2017). Dissection of additive, dominance, and imprinting effects for production and reproduction traits in Holstein cattle. *BMC Genomics* **18**: 425.

Kempthorne O (1954). The correlation between relatives in a random mating population. *Proc R Soc Lond B* **143**: 103–113.

Kinghorn B (1987). On computing strategies for mate allocation. *J Anim Breed Genet* **104**: 12–22.

Kocher TD, Lee WJ, Sobolewska H, Penman D, McAndrew B (1998). A genetic linkage map of a cichlid fish, the tilapia (Oreochromis niloticus). *Genetics* **148**: 1225–32.

Laipis PJ, Wilcox CJ, Hauswirth WW (1982). Nucleotide Sequence Variation in Mitochondrial Deoxyribonucleic Acid from Bovine Liver1. *J Dairy Sci* **65**: 1655–1662.

Lee B-YY, Lee W-JJ, Streelman JT, Carleton KL, Howe AE, Hulata G, *et al.* (2005). A second-generation genetic linkage map of tilapia (Oreochromis spp). *Genetics* **170**: 237–44.

Lee SH, van der Werf JHJ, Hayes BJ, Goddard ME, Visscher PM (2008). Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet* **4**: e1000231.

Legarra A (2016). Comparing estimates of genetic variance across different relationship models. *Theor Popul Biol* **107**: 26–30.

Leroi AM (2010). *Function and constraint in Aristotle and evolutionary theory*. na.

Li M-H, Yang H-H, Li M-R, Sun Y-L, Jiang X-L, Xie Q-P, *et al.* (2013). Antagonistic roles of Dmrt1 and Foxl2 in sex differentiation via estrogen production in tilapia as demonstrated by TALENs. *Endocrinology* **154**: 4814–4825.

Li M, Yang H, Zhao J, Fang L, Shi H, Li M, *et al.* (2014). Efficient and heritable gene targeting in tilapia by CRISPR/Cas9. *Genetics*: genetics-114.

Lien S, Koop BF, Sandve SR, Miller JR, Kent MP, Nome T, *et al.* (2016). The Atlantic salmon genome provides insights into rediploidization. *Nature* **533**: 200.

Liu Z (2017). *Genome-based Biotechnologies in Aquaculture (Draft Report)*.

Lopes MS, Bastiaansen JWM, Harlizius B, Knol EF, Bovenhuis H (2014). A Genome-Wide Association Study Reveals Dominance Effects on Number of Teats in Pigs (S Moore, Ed.). *PLoS One* **9**: e105867.

Lopes MS, Bastiaansen JWM, Janss L, Knol EF, Bovenhuis H (2015). Genomic prediction of growth in pigs based on a model including additive and dominance effects. *J Anim Breed Genet*.

Lynch M, Walsh B (1998). *Genetics and analysis of quantitative traits*. Sinauer Associates, Inc. Publishers, Sunderland, Massachusetts, US.

Mabry CC, Denniston JC, Nelson TL, Son CD (1963). Maternal phenylketonuria: a cause of mental retardation in children without the metabolic defect. *N Engl J Med* **269**: 1404–1408.

Mackay TFC (2014). Epistasis and quantitative traits: using model organisms to study gene–gene

interactions. *Nat Rev Genet* **15**: 22.

MacLeod IM, Hayes BJ, Goddard ME (2013). Will sequence SNP data improve the accuracy of genomic prediction in the presence of long term selection. In: *Proc Assoc Advmt Anim Breed Genet*, Vol 20, pp 215–219.

Marshall CT, Kjesbu OS, Yaragina NA, Solemdal P, Ulltang Ø (1998). Is spawner biomass a sensitive measure of the reproductive and recruitment potential of Northeast Arctic cod? *Can J Fish Aquat Sci* **55**: 1766–1783.

McGrath J, Solter D (1984). Completion of mouse embryogenesis requires both the maternal and paternal genomes. *Cell* **37**: 179–183.

Mendel G (1866). Versuche über Pflanzenhybriden. *Verhandlungen des naturforschenden Vereines Brunn 4 3* **44**.

Meuwissen THE, Goddard ME (2010). Accurate prediction of genetic values for complex traits by whole genome resequencing. *Genetics*.

Mishra SP, Taraphder S, Roy M, Sarkar U, Datta S, Saikhom R, *et al.* (2017). Breeding techniques to exploit non-additive gene action for improvement of Livestock. *Bull Env Pharmacol Life Sci* **6**: 126–134.

Misztal I (1997). Estimation of variance components with large-scale dominance models. *J Dairy Sci* **80**: 965–974.

Moore AJ, Brodie III ED, Wolf JB (1997). Interacting phenotypes and the evolutionary process: I. Direct and indirect genetic effects of social interactions. *Evolution (N Y)* **51**: 1352–1362.

Mor A, Avtalion RR (1990). Transfer of antibody activity from immunized mother to embryo in tilapias. *J Fish Biol* **37**: 249–255.

Moreno-Gonzalez J, Hallauer AR (1982). Combined S 2 and crossbred family selection in full-sib reciprocal recurrent selection. *Theor Appl Genet* **61**: 353–358.

Morgante F, Huang W, Maltecca C, Mackay TFC (2018). Effect of genetic architecture on the prediction accuracy of quantitative traits in samples of unrelated individuals. *Heredity (Edinb)* **120**: 500–514.

Mrode RA (2014). *Linear models for the prediction of animal breeding values*. Cabi.

Muir WM (2005). Incorporation of Competitive Effects in Plant or Animal Breeding Programs. *Genetics*.

Muñoz PR, Resende MDVMFR, Gezan SA, Resende MDVMFR, de Los Campos G, Kirst M, *et al.* (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics* **198**: 1759–68.

NCBI (2018). Assembly information by organism.

Neira R (2010). Breeding in aquaculture species: genetic improvement programs in developing countries. In: *Proceedings of the 9th World Congress on Genetics Applied to Livestock Production*, Vol 8, p .

Palaiokostas C, Bekaert M, Khan MGQ, Taggart JB, Gharbi K, McAndrew BJ, *et al.* (2013). Mapping and Validation of the Major Sex-Determining Region in Nile Tilapia (Oreochromis niloticus L.) Using RAD Sequencing (L Orban, Ed.). *PLoS One* **8**: e68389.

Pante MJR, Gjerde B, McMillan I, Misztal I (2002). Estimation of additive and dominance genetic variances for body weight at harvest in rainbow trout, Oncorhynchus mykiss. *Aquaculture* **204**: 383–392.

Piaskowski J, Hardner C, Cai L, Zhao Y, Iezzoni A, Peace C (2018). Genomic heritability estimates in sweet cherry reveal non-additive genetic variance is relevant for industry-prioritized traits. *BMC Genet* **19**: 23.

Poutous M, Vissac B, Calomiti M, Chambeyron M (1962). Recherche théorique des conditions de rentabilité maximum de l'épreuve de descendance des taureaux d'insémination artificielle. In: *Annales de zootechnie*, Vol 11, pp 233–256.

Powell JE, Henders AK, McRae AF, Kim J, Hemani G, Martin NG, *et al.* (2013). Congruence of additive and non-additive effects on gene expression estimated from pedigree and SNP data. *PLoS Genet* **9**: e1003502.

Pryce JE, Haile-Mariam M, Goddard ME, Hayes BJ (2014). Identification of genomic regions associated with inbreeding depression in Holstein and Jersey dairy cattle. *Genet Sel Evol* **46**: 71.

Puttaraksar N (2004). *GIFT technology manual: an aid to tilapia selective breeding*. WorldFish Center.

Raidan FSS, Porto-Neto LR, Li Y, Lehnert SA, Vitezica ZG, Reverter A (2018). Evaluation of non-additive effects in yearling weight of tropical beef cattle. *J Anim Sci*.

Ravi K, Vadez V, Isobe S, Mir RR, Guo Y, Nigam SN, *et al.* (2011). Identification of several small main-effect QTLs and a large number of epistatic QTLs for drought tolerance related traits in groundnut (Arachis hypogaea L.). *Theor Appl Genet* **122**: 1119–1132.

de Roos APW, Hayes BJ, Goddard ME (2009). Reliability of Genomic Predictions Across Multiple Populations. *Genetics* **183**: 1545 LP-1553.

Rothschild MF, Ollivier L (1987). Expectation of variance due to mitochondrial genes from several mating designs. *Genet Sel Evol* **19**: 171.

Roughsedge T, Brotherstone S, Visscher PM (2001). Bias and power in the estimation of a maternal family variance component in the presence of incomplete and incorrect pedigree information. *J Dairy Sci* **84**: 944–950.

Rye M, Mao ILL (1998). Nonadditive genetic effects and inbreeding depression for body weight in Atlantic salmon (Salmo salar L.). *Livest Prod Sci* **57**: 15–22.

Sae-Lim P, Gjerde B, Nielsen HM, Mulder H, Kause A (2016). A review of genotype-by-environment interaction and micro-environmental sensitivity in aquaculture species. *Rev Aquac* **8**: 369–393.

Sellier P (1976). The basis of crossbreeding in pigs; a review. *Livest Prod Sci* **3**: 203–226.

Shaw FH, Woolliams JA (1999). Variance component analysis of skin and weight data for sheep subjected to rapid inbreeding. *Genet Sel Evol* **31**: 1.

Smith C (1964). The use of specialised sire and dam lines in selection for meat production. *Anim Sci* **6**: 337–344.

Smith CC, Fretwell SD (1974). The optimal balance between size and number of offspring. *Am Nat* **108**: 499–506.

Smith SP, Mäki-Tanila A (1990). Genotypic covariance matrices and their inverses for models allowing

dominance and inbreeding. *Genet Sel Evol* **22**: 65.

Solemdal P (1997). Maternal effects—a link between the past and the future. *J Sea Res* **37**: 213–227.

Sonesson AK, Meuwissen THE (2009). Testing strategies for genomic selection in aquaculture breeding programs. *Genet Sel Evol* **41**: 37.

Speed D, Balding DJ (2015). Relatedness in the post-genomic era: is it still useful? *Nat Rev Genet* **16**: 33.

Su G, Christensen OF, Ostersen T, Henryon M, Lund MS (2012). Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One* **7**: e45293.

Surani MAH, Barton SC, Norris ML (1984). Development of reconstituted mouse eggs suggests imprinting of the genome during gametogenesis. *Nature* **308**: 548.

Swain P, Nayak SK (2009). Role of maternally derived immunity in fish. *Fish Shellfish Immunol* **27**: 89–99.

Toro MA, Mäki-Tanila A (2018). Some intriguing questions on Fisher's ideas about dominance. *J Anim Breed Genet* **135**: 149–150.

Toro MA, Varona L (2010). A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol* **42**.

Vandeputte M, Quillet E, Chevassus B (2002). Early development and survival in brown trout (Salmo trutta fario L.): indirect effects of selection for growth rate and estimation of genetic parameters. *Aquaculture* **204**: 435–445.

VanRaden PM (2008). Efficient Methods to Compute Genomic Predictions. *J Dairy Sci* **91**: 4414–4423.

Varona L, Legarra A, Toro MA, Vitezica ZG (2018). Non-additive Effects in Genomic Selection. *Front Genet* **9**: 78.

Vinkhuyzen AAE, Wray NR, Yang J, Goddard ME, Visscher PM (2013). Estimation and partition of heritability in human populations using whole-genome analysis methods. *Annu Rev Genet* **47**: 75–95.

Vitezica ZG, Legarra A, Toro MA, Varona L (2017). Orthogonal Estimates of Variances for Additive, Dominance, and Epistatic Effects in Populations. *Genetics* **206**: 1297–1307.

Vitezica ZG, Varona L, Elsen J-M, Misztal I, Herring W, Legarra A (2016). Genomic BLUP including additive and dominant variation in purebreds and F1 crossbreds, with an application in pigs. *Genet Sel Evol* **48**: 6.

Vitezica ZG, Varona L, Legarra A (2013). On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* **195**: 1223–30.

Walsh B (2005). The struggle to exploit non-additive variation. *Aust J Agric Res* **56**: 873–881.

Wang CC-H, Li SS-F, Xiang S-PS, Wang J, Liu ZZ-G, Pang Z-YZ, *et al.* (2006). Additive, dominance genetic effects for growth-related traits in common carp, Cyprinus carpio L. *Aquac Res* **37**: 1481–1486.

Watanabe S, Watanabe Y (2002). Relationship between male size and newborn size in the seaweed pipefish, Syngnathus schlegeli. *Environ Biol Fishes* **65**: 319–325.

Wei M, van der Werf JHJ (1994). Maximizing genetic response in crossbreds using both purebred and crossbred information. *Anim Sci* **59**: 401–413.

Weimin M (2017). Aquaculture production and trade trends: carp, tilapia and shrimp. In: *FMM/RAS/298: Strengthening capacities, policies and national action plans on prudent and responsible use of antimicrobials in fisheries Final Workshop in cooperation with AVA Singapore and INFOFISH*, FAO: Singapore, p .

Weller JI (1994). *Economic aspects of animal breeding*. Springer Science & Business Media.

Wiener G, Lee GJ, Woolliams JA (1992a). Effects of rapid inbreeding and of crossing of inbred lines on the body weight growth of sheep. *Anim Sci* **55**: 89–99.

Wiener G, Lee GJ, Woolliams JA (1992b). Effects of rapid inbreeding and of crossing inbred lines on the growth of linear body dimensions of sheep. *Anim Sci* **55**: 101–114.

Winkelman AM, Peterson RG (1994a). Genetic parameters (heritabilities, dominance ratios and genetic correlations) for body weight and length of chinook salmon after 9 and 22 months of saltwater rearing. *Aquaculture* **125**: 31–36.

Winkelman AM, Peterson RG (1994b). Heritabilities, dominance variation, common environmental effects and genotype by environment interactions for weight and length in chinook salmon. *Aquaculture* **125**: 17–30.

Wittenburg D, Melzer N, Reinsch N (2011). Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers. *BMC Genet* **12**: 74.

Woolliams JA, Wiener G (1980). The effects of breed, crossbreeding and other factors on variation in fleece and skin traits. *Anim Sci* **30**: 417–429.

Worldfish (2016). A new research project on genetic improvement.

Wright S (1921). Systems of mating. I. The biometric relations between parent and offspring. *Genetics* **6**: 111.

Xiang T, Christensen OF, Vitezica ZG, Legarra A (2016). Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genet Sel Evol* **48**: 92.

Yang J, Benyamin B, McEvoy BP, Gordon S, Henders AK, Nyholt DR, *et al.* (2010). Common SNPs explain a large proportion of the heritability for human height. *Nat Genet* **42**: 565–569.

Yustiati A, Bangkit I, Zidni I (2018). Masculinization of nile tilapia (Oreochromis niloticus) using extract of bull testes. In: *IOP Conference Series: Earth and Environmental Science*, IOP Publishing Vol 139, p 12008.

ISBN: 978-82-575-1567-6
ISSN: 1894-6402

Norwegian University
of Life Sciences

Postboks 5003
NO-1432 Ås, Norway
+47 67 23 00 00
www.nmbu.no