



Norwegian University
of Life Sciences

Master's Thesis 2019 60 ECTS

Department of Animal and Aquacultural Sciences (IHA) Faculty of Biosciences

Estimation of genetic parameters and strain genetic effects in *Oreochromis mossambicus*

Koushik Das

Masters of Science - Animal Breeding and Genetics

**Estimation of genetic parameters and strain
genetic effects in *Oreochromis mossambicus***

By

Koushik Das

Supervisor

Professor Hans Magnus Gjøen

Department of Animal and Aquacultural Sciences

Faculty of Biosciences

Norwegian University of Life Sciences

August 2019

Abstract

A breeding program for *Oreochromis mossambicus* has been established by CEPAQ (Centro de Pesquisa em Aquacultura, translated as The Research Centre in Aquaculture) and is intended to develop the local fish farming industries in Mozambique. As much of the area do have high salinity, *O. mossambicus* is considered suitable for cultivation as they have the capacity to grow even in high estuarine condition. To genetically improve and produce quality fingerlings of *O. mossambicus*, the base population was created by using the best performing pure strain and strain combination as parents in the breeding program.

The aim of this study was thus to estimate genetic parameters in three strains of Mozambique tilapia (*Oreochromis mossambicus*), this includes the degree of strain additive, reciprocal and heterosis effects for different traits. A partial factorial mating design with reciprocal crosses and parentage assignment using Genotype-By-Sequencing (GBS) data were used to estimate the proportional effect of additive (h^2) dominance (d^2) and maternal effects (m^2) for six different traits in Mozambique Tilapia. The study includes observations of 1119 individuals from 300 fullsib families. Two different nested statistical models, an Additive-Dominance-Maternal (ADM) model and a Sir-Fullsib-Maternal (SFM) model, including their reduced variants, were used to test the significance of each effect in the full model.

The estimates of heritability reported by the ADM models and SFM models were almost similar for most traits; e.g. the heritability of Harvest Body Weight (HBW) for the ADM model and SFM model were 0.13 and 0.14 respectively. A large amount of dominance was observed for some of the traits, e.g. it was 0.45 for harvest body weight when using the SFM model.

A separate general linear mixed model was also used to describe the strain additive, reciprocal and heterosis effects for each trait. For harvest body weight, crossing between S2 and S3 strain yielded positive heterosis effect. The reciprocal effects were non-significant for all the crosses and the additive strain effect was only significant for the S1 strain. Crossing S2 and S3 strain seems to be the best way forward to improve the breeding population, either by a pure-breeding or a cross-breeding scheme.

Acknowledgement

First and foremost, I would like to express my deepest appreciation to my supervisor, Prof. Hans Magnus Gjøen at the Department of Animal and Aquacultural Sciences, here at the Norwegian University of Life Sciences. The door of Prof. Gjøen office was always open whenever I needed his assistance; who in my view has a right attitude and a good intellect to convey the spirit of adventure to me regarding research.

Besides my supervisor, I would also like to thank Dr Rajesh Joshi, Senior Researcher, GenoMar Genetics for his motivation, insightful comments and endless support.

My sincere thanks also go to Mr. Guillermo Torres Arjona and other team members working at CEPAQ; they have struggled hard to set up the genetic enhancement program of *O. mossambicus* and done a terrific job recording good and valid data.

Finally, I would also like to thank my family: my parents Kamal Prasad Das and Nilima Das, for supporting me throughout my life.

Contents

Abstract	1
Acknowledgement	2
Introduction	7
Selective breeding	8
Breeding goal.....	8
Establishment of a base population	8
Breeding strategies.....	9
Mating design.....	10
Using the relationship matrices in the Animal Model	11
Genotyping-by-sequencing	12
Objectives of the study.....	13
Materials and methods	13
Experimental facilities	13
Collection of wild breeders of <i>O. mossambicus</i>	14
Experimental design and rearing of fish	14
Collection of data	15
DNA sampling.....	15
Genotyping.....	15
Parentage assignment.....	16
Data recording and correction	17
Statistical analysis.....	17
SFM model	17
ADM model	18
SRH model	19
Results.....	21
Descriptive statistics.....	21
ADM and SFM models	22
SRH Model.....	25
Discussion.....	27
Additive and non-additive genetic strain effects.....	27
Heritability.....	27
Non-additive genetic variance components	28

Best model fitted	28
Population subdivision	29
Conclusion.....	29
References	31
Supplement 1	35
Supplement 2	37
Supplement 3	38
Supplement 4	39
Supplement 5	40
Supplement 6	41
Supplement 7	42
Supplement 8	43

List of tables

Table 1- Strain distribution over the different location in Mozambique	14
Table 2- . Descriptive statistics of Phenotypic measurements or trait	21
Table 3- . Mean and CV for each strain groups among all traits	22
Table 4- Goodness of fit	23
Table 5- Heritability, dominance ratio, and maternal Ratio \pm SE for the best fit models.	25
Table 6- Estimates of additive genetic, reciprocal and total heterosis effects of all phenotypic measurements	26

List of figures

- Figure 1. The heatmap based on estimated genetic relatedness among the wild breeders. 16
- Figure 2. Segregation of the phenotypic variance of all traits into additive genetic (h^2), dominance (d^2), maternal (m^2), and error component (e^2) 24

Introduction

The production of farmed aquatic species has increased by a factor of four between 1995 and 2015, from 12 to 51 million tons. New intensive production systems for shrimps, tilapias, carps, and salmonids have largely contributed to this increase. The production of Tilapia and other freshwater species are expected to represent around 60 per cent of the total world aquaculture production in 2030 (FAO, 2018). Among the many Tilapia species, three are particularly popular and commercialised; Nile tilapia (*Oreochromis niloticus*), Mozambique Tilapia (*Oreochromis mossambicus*) and Blue Tilapia (*Oreochromis aureus*) (Alda Ma. Salia 2008). One special feature of *O. mossambicus* is that they can survive in a wide range of salinities because of their osmoregulatory capabilities, as seen by high levels of organic osmolytes and cortisol as well as prolactin hormones in blood plasma (Fiess et al., 2007; Whitfield et al., 2006). Unlike other freshwater fish, *O. mossambicus* can spawn in hypersaline conditions, whereas the other fish species stop reproducing in brackish water conditions. For example, the presence of a high number of juvenile *O. mossambicus* in St. Lucia indicates successful spawning of the fish in high salty conditions (Cyrus et al., 2005; Whitfield et al., 2006). In a comparative study with Nile Tilapia, *O. mossambicus* had better salinity acclimation, both *in vivo* and *in vitro*. (Yamaguchi et al., 2018).

Hybridisation of local species by invasive species is generally considered as a common problem when developing commercial strains of that species. The genetic introgression of *O. mossambicus* by two invasive congeneric species (*Oreochromis niloticus* and *Oreochromis andersonii*) has already been reported in some regions of Mozambique.

On the other hand, *O. mossambicus* can be a potential candidate for sustainable aquaculture development, due to its unique properties, i.e. outstanding euryhalinity, and its high market values, especially the reddish-orange mutant. *O. mossambicus* has thus been identified as a good candidate for a genetic improvement program, both to protect the genetic variation and culture salinity-resistant pure breed for future breeding (Firmat et al., 2013). Poor-quality broodstock is constraining the commercial tilapia production in many areas of Africa and Asia (Modadugu & O. Acosta, 2004), and therefore it becomes crucial to develop genetically improved broodstock for the commercial cultivation of *O. mossambicus*.

Selective breeding

Profitable aquaculture production is largely dependent on the access to genetically improved broodstock, and selective breeding is thus currently used as an important tool to improve industrial aquaculture production (Gjedrem et al., 2012). The selective breeding program is mainly defined by the selection intensity, the number of sib families and the mating scheme. The program should use a proper selection strategy with the sib-families originating from a pre-planned mating design and a sufficient number of breeding candidates per selection round (Bentsen & Gjerde, 1994; Gjerde, 2005).

First, a solid base population should be generated by collecting breeders with a high genetic potential. The base population can be collected from the wild populations or domestic populations, or from both. Later, to achieve long-term selection response, the additive genetic variability of the desired traits should be maintained for every generation by control of inbreeding. The inbreeding is controlled by restricting the amount of selected candidates from each family or by applying the optimal contribution procedure in every generation (Gjedrem et al., 1991; Gjøen & Bentsen, 1997; Nielsen et al. 2011).

Breeding goal

In general, animal breeding programs emphasise the gradual genetic improvement of the given traits in a breeding program, which is dependent on a long-term and meticulous process. The desired traits should also have large genetic variance, must be heritable and measurable with a reasonable cost (Gjedrem, 2005). The main breeding goal normally includes growth rate, followed by disease resistance and carcass quality. For example, in a Chilean genetic improvement program of salmonids, the main breeding goal included resistance to a specific disease alongside the growth rate (Lhorente et al., 2019).

Establishment of a base population

The base population needs wide genetic variability, which often leads managers to making a synthetic F1 population, i.e. a combination of different subpopulations that has considerable genetic distance among them. The heterogenetic effects of locus generally increase by mixing different fish stocks that have divergent gene pools. As a result, the additive genetic and non-additive genetic variation can be maintained to achieve genetic gains in subsequent generations (Gjedrem, 2005). In the Genetically Improved Farmed Tilapia (GIFT) breeding program in the

Philippines, four wild Nile Tilapia strains and four domestic strains (Eknath et al., 1993) were sampled to form the base population. Later, the F1 generation was created from the base population by applying low selection intensity and diallel cross design, which again was followed by large selection responses in a later generation (Bentsen et al., 1998).

However, at the time of creating the base population, the genetic-relationship information among strain groups or individuals within each strain may not be available. In this case, the best strategy is basically to use as many strains as possible or feasible. To exploit potential additive genetic gain in the next generations, fish should be sampled from at least four subpopulations. Also, equal contribution of each strain should be ensured in the base population, i.e. including an equal number of individuals from each strain, in order to maintain a lasting genetic variability (Fernández et al., 2014; Holtsmark et al., 2006).

Breeding strategies

For any breeding program, it is necessary to apply a proper breeding strategy to genetically improve the desired traits in the breeding population. Pure breeding and crossbreeding are popular breeding strategies and figuring out the proper method relies on the genetic variation of the desired trait, either originating from additive and/or non-additive genetic sources. Pure breeding is a common breeding strategy and exploits mainly the additive genetic effects. On the other hand, crossbreeding is performed between breeds or lines, and generally increases the heterozygosity or heterogenetic effect of locus of the desired trait (Gjedrem, 2005). Maluwa & Gjerde (2006b) suggested that a proper breeding strategy is necessary to prevent the loss of genetic resources of the founder population. To select an optimum breeding strategy, type and amount of the genetic variation of the desired traits should be known (William & Pollak, 1985).

Therefore, understanding the genetic resources of the trait is helpful when setting up the breeding objectives and selecting the best suited breeders, in order to maximise the genetic gain. Based on the genetic and non-genetic sources, the total phenotypic variance can be partitioned into some casual factors, such as genetic, environment and interaction between genetic and environment. The genetic part of the variance is further subdivided into an additive portion that can be estimated as the average effects of the genes/alleles, the dominance portion caused by dominance interaction within locus, and a portion originated from interaction among locus, the epistatic effect. The estimated breeding value (EBV), a prediction of the additive effect of the gene, is subject to the

selection decision, although the true genetic potential of an animal is not possible to estimate in practice (Fisher, 1919; Joshi, 2018; Visscher & Bruce Walsh, 2017). Under the conditions described by Cockerham (1954), the partitioning of the genetic variance into genetic and non-genetic sources is orthogonal, therefore, the inclusion of the terms called dominance and epistatic effects should more accurately predict the EBVs (Muñoz et al., 2014). But the presence of the Hardy-Weinberg Equilibrium and linkage disequilibrium alters the orthogonal properties of the estimates of those variance components (Joshi, 2018).

The non-additive variance can be as large as the additive genetic variance, or in some cases even larger, for different traits and could cause a significant change in the genetic evaluation results (Palucci et al., 2007). Although this non-additive genetic variance is not transmissible to the offspring, the genetic variances estimated may show significant heterosis effects. This heterozygosity, largely derived from the dominance and epistasis effects, indicates the genetic potential of the parents' group, and could be considered for commercial evaluation (Joshi, 2018).

The heterosis, for which the performance of offspring is superior to the parental strains or lines, is generally observed when two separate inbred lines or pure strain groups are crossed. Therefore, the crossbreeding program acts as a supplement to the additive genetic improvement resulting from the pure breeding program. Different crossbreeding strategies can be used to establish an efficient breeding program by exploiting heterosis (Fjalestad, 2005), and the crossbreeding program has a role in developing the whole breeding program by selecting the best-purebred parent based on crossbred performance. Crossbreeding thus has a potential for greater total selection response to selection. In addition, crossbreeding helps to protect the genetic assets of the breeding company and increases the sustainability of the company in the market without decreasing the additive genetic performance (Joshi, 2018).

Mating design

As a part of the breeding program, the mating design should be chosen carefully in order to improve the genetic resources in subsequent generations (Gjedrem & Robinson, 2014). Different mating designs have different capabilities of preserving the genetic variation in a population. Therefore, the design is important both for short-term and long-term genetic gain in the breeding program. When applying modern selection schemes, it is also required to keep the identification of each individual to implement a specific mating design (Dupont-Nivet et al., 2006). In fish breeding,

physical tagging will require that the fish has reached a certain size, which requires the keeping of families separate in a multi-tank facility until fish can be tagged, typically until they are 10-15 grams. This will normally increase the common environmental effects (Gjerde, 2005). Alternatively, genetic markers can be used to identify fish and to assign the parents (Dupont-Nivet et al., 2006; Gjerde, 2005). By applying a proper identification system, many sib families can be achieved, and the desired mating design can be easily implemented.

In the simplest mating design, called single pair mating design, only one male is used to fertilise one female, and the design can in some cases be implemented if the non-additive genetic effects and fullsib effects are low (Gjerde, 2005). A more commonly applied system is the nested mating design, whereby one sire is mated with two to three dams. Pante et al. (2002) found significant amount of dominance variance for harvest bodyweight in rainbow trout when applying this design, but the dominance variance was confounded with common environmental effects.

To better dissect these effects, the factorial mating design is proposed (Berg & Henryon, 1998), where each male is mated with several females and the eggs of one female are either partitioned into groups which are subsequently fertilised by different males or being mated subsequently to different males. As a result, several fullsibs groups, alongside maternal and paternal half-sibs groups, are being created (Gjerde, 2005). This design can reduce the correlation of EBVs among parents, resulting in a lower chance of selecting individuals from the same fullsib groups and reduced inbreeding (Sørensen et al., 2005).

Yet another mating system, applied by Joshi et al. (2018), is to apply a full factorial mating design that includes reciprocal crosses of for instance two lines. This setup can even better separate maternal and non-additive genetic effects. They showed that when maternal and dominance effects were also estimated, the ranking of the candidates based on their EBVs changed, which indicates that these non-additive genetic effects may influence the selection significantly. For practical reasons, as an alternative of the full factorial mating design, a partial factorial mating design can be implemented in the breeding program because it will still be able to dissect out most of the non-additive and common environmental effects (Dupont-Nivet et al., 2006).

Using the relationship matrices in the Animal Model

Pedigree-based relationship matrices can be utilised in linear mixed animal models to estimate the breeding values and to estimate the genetic as well as phenotypic parameters. The complete

pedigree-based relationship matrix is necessary to predict the unbiased genetic components (Kennedy et al., 1988; Kolstad, 2005). For example, in several studies (Raidan et al., 2018); (Muñoz et al., 2014) and (Su et al., 2012), the relationship matrices have been used to separate the total genetic variance components into additive and non-additive genetic parts.

As compared to pedigree-based information, the marker-based information seems to be more effective in separating the additive and non-additive variance components, where the dense panel of SNPs are used to define the genomic information (Muñoz et al., 2014). Also, Joshi (2018) found that by using pedigree-based relationship matrices, the source of non-additive genetic variation was falsely concluded as dominance. He also showed that the genomic relationship matrices described a major portion of the non-additive variance caused by epistasis in that population. But using the best genomic prediction methods did still not separate the non-additive genetic variances completely from the additive one, at least not in an orthogonal manner. Therefore, despite the fact the non-additive genetic effects improved the accuracy for predicting the EBVs, a serious weakness still remain in that the partitioning of additive and non-additive variance is not completely orthogonal, even when using SNP markers to construct more accurate genetic relationship matrices (Raidan et al., 2018).

Genotyping-by-sequencing

Estimating the relatedness of the individuals based on marker information can be a very efficient way to infer the population substructure and to estimate the breeding values. This marker-based realised relationship gives a more accurate estimation of the genetic variants between two closely related animals, which can also be used to define the unrecorded pedigree and increase the genetic merit (Dodds et al., 2015; Makgahlela et al., 2013). But the high-density SNPs are not developed for all aquatic species, so, in the absence of reference genome, genotyping-by-sequencing (GBS) can be applied to get unbiased relatedness values, and many broad-scale aquaculture breeding programs have thus applied GBS as a DNA-tool in order to reduce the costs. Finally, GBS techniques can also be applied in several research fields like genetic diversity, population structure, association analysis for economically important traits and genomic selection in aquaculture breeding (Robledo et al., 2018).

Objectives of the study

The Centro de Pesquisa em Aquaculture (CEPAQ), translated to English as The Research Centre in Aquaculture, is established to develop the Tilapia aquaculture industry in Mozambique. *O. mossambicus* is a potential candidate to boost up the local fish farming industry because of the species' high growing capacity in the high estuarine condition of Mozambique. It was thus decided to establish a composite base population, using best performing pure strain and strain combinations as a base population for commercial production of quality fingerlings and for running a long-term selection program for *O. mossambicus*. In this study, the harvest body weight and other morphometric traits are investigated.

Therefore, the main goal of this study is to describe the genetic component of the base population at CEPAQ and to estimate the effect of crossbreeding among the strains of *O. mossambicus*. The study has the following sub-goals:

- Estimate genetic and phenotypic parameters of the founder population produced in a partial factorial mating design among different strains of *O. mossambicus*.
- Estimate the genetic strain effects i.e. the strain additive genetic effect, reciprocal cross effect, heterosis effect of the different strains and strain combinations tested.

Materials and methods

Experimental facilities

CEPAQ is established in the district of Chokwe which is in one of the southern provinces of Mozambique, Gaza. The Research Centre has an area of approximately 10 500 ha and the land or soil is known for high salinity levels but still suitable for Tilapia farming. The facility is subdivided into three areas; the genetic enhancement area, hatchery area and, grow-out area. The research centre at the genetic enhancement area is dedicated to the genetic enhancement of the *O. mossambicus*. The Hatchery is developed for producing sex-reversed fingerlings and the grow-out area is aimed to develop the protocols and a training facility for the full production cycle.

Collection of wild breeders of *O. mossambicus*

The genetic enhancement program, CEPAQ was responsible for the collection of the pure breed of *O. mossambicus* from the wild environment, which was a very challenging task. But the sampling of the wild breeders was done efficiently by Arjona (2018) and his team members. Different strains of *O. mossambicus* were collected from 12 different sites, clustering into five different main locations, first defined as strains, of the southern and central regions of Mozambique, from October 2016 to June 2017. Later, fin samples were sent to AgResearch Limited, New Zealand. The following table shows the five different strains of *O. mossambicus* based on the different locations of Mozambique:

Table 1. Strain distribution over the different location in Mozambique (Arjona, 2018)

Strain	Location	Total number of fish
A	Sotiva	346
B	Bons Sinais	594
C	Catuane	219
D	Marrangua	538
E	Govuro, Govuro 3, Ximite, Makuri	357

Experimental design and rearing of fish

Total 921 wild breeders were selected as breeding candidates from the wild stock of *O. mossambicus*, which were based on some conditions: the fish were in the proper size for reproduction (>60g) and had optimal health status. Later the selected population was DNA-sampled and PIT-Tagged, after being anaesthetised using a solution of clove oil in ethanol.

A partial factorial mating design with reciprocal crosses was used to test the different strains. From this, a total of 25 different genotypes were created, including 20 crossbreds and 5 purebreds. To utilise manpower and facilities to the maximum, the mating was organised into 10 successive batches, each batch consisting of 50 hapas, where 2 replicas of each line were represented. Each pair could stay for 2 to 3 weeks in the hapa for successful spawning. All eggs were naturally incubated by the female alone in the hapa.

From every family, 25 fry were collected and transferred into rearing hapas. After collecting all fry in their respective batch, they were transferred into a grow-out pond. The grow-out period of each batch was about 100 days. In total, around 10347 fingerlings were stocked in grow-out ponds enriched with inorganic fertilisers for growing.

Collection of data

After grow-out, the fish were collected by draining out the pond and then gathered by a seine net. The traits that were recorded were as follows: harvest body weight HBW (g), total length, TL (cm), standard length, SL (cm), head length, HL (cm), body height, BH (cm), and body width, BW (cm).

DNA sampling

The DNA samples were collected from the caudal fins of the fish. The samples were preserved in small tubes with 96 trays filled with pure ethanol and then stored in a freezer.

Genotyping

The DNA samples including the wild breeders were sent to AgResearch Limited, New Zealand and Genotyped by sequencing (GBS). The genomic relationship matrix (G-matrix) was also provided by them which was constructed based on the method described by Dodds et al. (2015). Initially, 96 samples from the wild fish that were used as parents for the F1 generation, were genotyped to establish a suitable set of SNPs to use. In the final genotyping, 36542 SNPs were used with a mean sample depth 3.56 and call rate 0.55. The genotyping results can be visualised as a heatmap, as shown in Figure 1. As can be seen there, the parents could better be grouped in 3, rather than 5, genetic groups; indicated by the darker colour (red) in the connection between related individuals. Less related individuals are marked with a light colour (yellow). Later, the five strains were thus grouped into three genetic groups or meta-strains, S1-S3, so that fish from Bons Sinais and Govuro are termed S1, from Marrangua S2 and from Sotiva and Catuane S3.

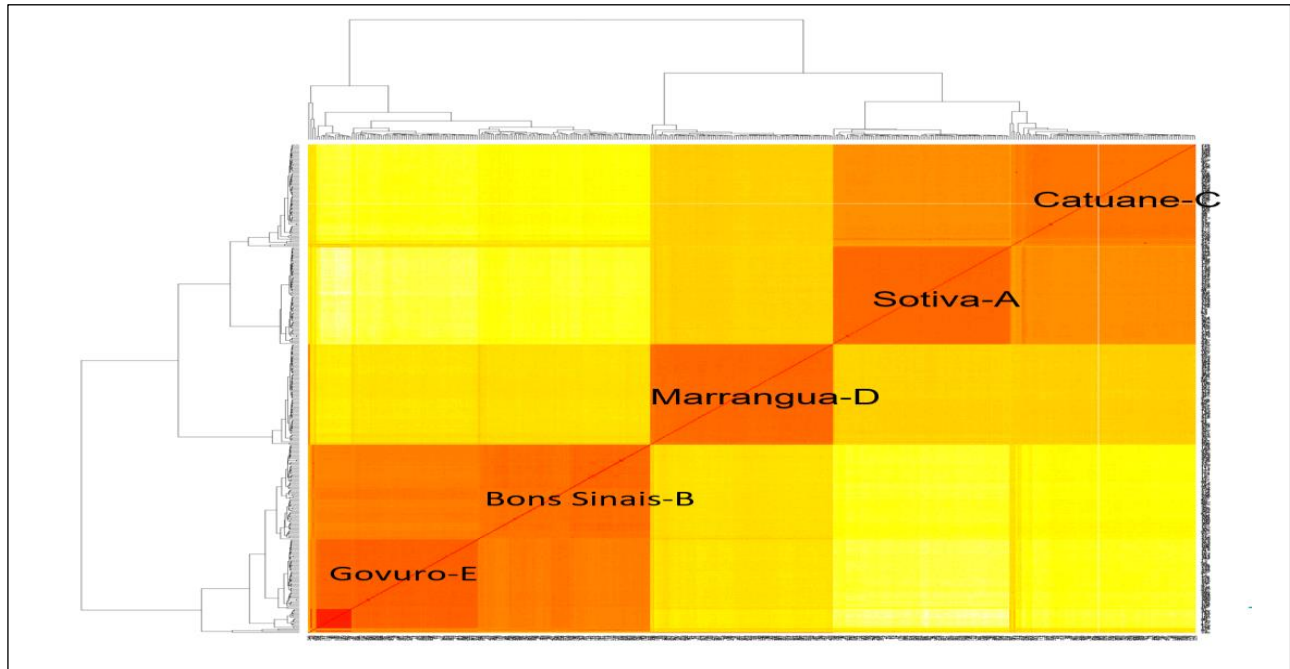


Figure 1. The heatmap based on estimated genetic relatedness among the wild breeders.

Parentage assignment

Based on the genomic relationship matrix, estimates of pairwise relationship, i.e. kinship coefficients, were used to create a list of potential parents for each offspring. The parentage assignments were done by considering the highest relationship coefficient values between offspring and potential parents. The potential parents for each offspring were sorted according to their relatedness value, so that, the sire or dam having the highest relatedness value was placed on top of the potential parent list. If that first-positioned parent in the list was not valid by matching on the recorded mating list, then the second top positioned parent was considered as the assigned parent, as with the previous procedure. For each iteration, targeting to assign true parent, the whole dataset was separated into assigned and non-assigned offspring. The whole parentage assignment was completed by tailormade scripts written in R and the total procedure is described in figure 2, supplement 3. However, this method left some offspring unassigned, and an alternative manually assignment procedure was performed for the remaining offspring. Figure-1 Supplement 3 describes the manual procedures for assigning parents when batch ID of individuals were available, whereas figure-1 Supplement 4, describes the manual procedures when the batch ID of individuals were not available.

Data recording and correction

In total, 2202 offspring were genotyped and phenotyped, but only about half of this, 1119 offspring, were used for statistical analysis since only male offspring were used in the genetic analysis. This was caused by the many females that had reached early sexual maturation, as well as spawned, at the time of harvest and thus would not serve as a true representation of the traits measured. Before analysis, all type of data records was evaluated or checked for abnormality, the measurements which significantly distorted from the sub-group means were marked as well as corrected by re-checking. Also, while performing the final statistical analysis in ASReml (see below), the outliers noted by the program were checked for validity and considered as missing values if it clearly negatively deviated from the mean value of a specific batch, as this may indicate that it was not growing properly, which could be due to an infection or distress, making it unsuited to measure the trait of interest. No significant outliers were detected for TL and SL, whereas, the four other traits had some very few of them. The five local strains described earlier (A, B, C, D, and E) were recoded into three strains denoted as S1, S2 and S3, as described above.

Statistical analysis

All the statistical analyses were performed using ASReml V4.1 (Butler et al., 2009), for the six traits harvest body weight HBW (g), total length, TL (cm), standard length, SL (cm), head length, HL (cm), body height, BH (cm), and body width, BW (cm). Three different models were used for the statistical analysis: A Sire, Maternal and Fullsib model (SFM) and an Additive, Dominance and Maternal model (ADM) and one were Strain additive genetic effects, Reciprocal effects, and Heterosis effects (SRH) were estimated:

SFM model

The partial factorial mating design used allows the separation of the total variances into several components, which in the Sire, Maternal, and Fullsib model (SFM) model are:

$$y = Xb + Z_s S + Z_m M + Z_f F + e$$

where y is the vector of phenotypes; b is the vector of fixed effects, the 10 batches ; S is a vector of random Sire effects; M is a vector of random Dam effects, F is a vector of random fullsib effects; X , Z_s , Z_m , and Z_f are the corresponding design matrices for the fixed and random effects. The

variance of Sire (σ^2_s) and variance of Dam (σ^2_M) were constrained at the same level in models S and SF to obtain appropriate additive genetic contributions.

$$\begin{bmatrix} S \\ D \\ F \\ e \end{bmatrix} = \begin{bmatrix} IV_{Sire} & 0 & 0 & 0 \\ 0 & IV_{Dam} & 0 & 0 \\ 0 & 0 & IV_{Fsib} & 0 \\ 0 & 0 & 0 & IV_e \end{bmatrix}$$

The variance of fullsib, σ^2_F was restrained to 0 in S, and unrestrained in SF. Sire variance and dam variance in the model SM and SFM were unconstrained, whereas σ^2_F was restrained to be 0 in SM and unrestrained in SFM. Phenotypic variance, σ^2_P was calculated as the sum of the variance of Sire (σ^2_s), the variance of Dam (σ^2_M), the variance of Fullsib (σ^2_F) and residual variance (σ^2_e). The additive genetic variance, σ^2_A was calculated as $4 \times \sigma^2_s$ and the dominance variance, σ^2_D was calculated as $4 \times \sigma^2_F$. Heritability, Dominance ratio, maternal ratio were calculated as: $h^2 = \frac{4 \times \sigma^2_s}{\sigma^2_P}$, $d^2 = \frac{4 \times \sigma^2_F}{\sigma^2_P}$, $m^2 = \frac{(\sigma^2_D - \sigma^2_s)}{\sigma^2_P}$ (Joshi et al., 2018).

ADM model

As an alternative to the SFM model, an animal model with Additive, Dominance, and Maternal effects (ADM) were also applied:

$$y = Xb + Z_1A + Z_2D + Z_3M + e$$

where y is the vector of phenotypic records for all traits, b is the vector of fixed effects as in the SFM model, A is the vector of individual random additive genetic effects; D is the vector of random dominance effects; M is the vector of random maternal effects; and e is the random error effects. X, Z₁, Z₂, and Z₃ are the design matrices corresponding to the fixed, additive, dominance, and maternal effects. Vectors A and D are fitted for everyone in the pedigree, but the maternal effects, M, are pertaining to each fullsib family.

In the ADM model, σ^2_{add} , σ^2_{dom} , σ^2_{mat} , and σ^2_e are additive genetic variance, dominance variance, maternal variance, and error variance respectively;

$$\begin{bmatrix} A \\ D \\ M \\ e \end{bmatrix} = \begin{bmatrix} A\sigma_{add}^2 & 0 & 0 & 0 \\ 0 & D\sigma_{dom}^2 & 0 & 0 \\ 0 & 0 & I\sigma_{mat}^2 & 0 \\ 0 & 0 & 0 & I\sigma_e^2 \end{bmatrix}$$

A is the additive relationship matrix, D is the Dominance relationship matrix, and I is the identity matrix with an appropriate size of the M vector. The phenotypic variance is calculated as $\sigma_p^2 = \sigma_{add}^2 + \sigma_{dom}^2 + \sigma_{mat}^2 + \sigma_e^2$. The heritability (h^2), dominance ratio (d^2), and maternal ratio (m^2) are expressed as the relative to the phenotypic variance of additive genetic variance, dominance variance, and maternal variance respectively. To fit the dominance relationship matrix in the ADM model, an inverse of the dominance matrix is calculated by using the R package “nadiv” as described in Wolak (2012). Models with the dominance, maternal, and both effects removed were also tested.

Goodness of fit for the various models 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 χ_1^2 , i.e., 2.71 (Joshi et al, 2018).

SRH model

The strain additive genetic effect, reciprocal cross effect, and heterosis effects for all traits were estimated using a univariate model termed Strain, Reciprocal, and Heterosis (SRH) model (Workagegn, 2019):

$$y_{ijklk} = \mu + Batch_k + \sum b_{ai}A_i + \sum b_{ri}R_i + \sum b_{hj}h_j + A_0 + e_{ijklk}$$

where y_{ijklk} is the record for each trait, on the l^{th} individual of the cross between the i^{th} and j^{th} strain groups reared in the k^{th} batch; μ is the overall mean; $Batch_k$ is the fixed effect of the k^{th} batch ($k=1-10$); b_{ai} is the regression coefficient of the additive genetic effect if the genes originating from the i^{th} strain ($i = 1-3$), A_i is the proportion of genes in the l^{th} individual originating from the i^{th} strain ($A_i= 0, 0.5$ or 1 , and $\sum A_i = 1$); b_{ri} is the regression coefficient of the general reciprocal effect for the i^{th} strain; R_i is the proportion of genes of the dam of the l^{th} individual originating from i^{th} strain ($R_i= 0$ or 1 and $\sum R_i = 1$); b_{hij} is the regression coefficient of the mean heterosis effect of both reciprocals of the j^{th} cross between two different strains ($j = 1-3$), h_j is the proportion of the

total heterosis effect of the j^{th} cross-strain group expressed in the l^{th} individual, ($h_j = 0$ or $1, \sum h_j = 1$ for cross strains or $\sum h_j = 0$ for pure strains), A_0 is the random additive genetic effect of the l^{th} individual fish, $A_0 \sim N(0, \sigma^2 A)$ where A is the additive genetic relationship matrix among all fish population and $\sigma^2 A$ is the additive genetic variance, and e_{ijkl} is the random residual error of the l^{th} individual. The details of the regression coefficient of the strain genetic effects; strain additive genetic effects, strain heterosis effects and strain reciprocal effects are shown in table-1 supplementary 7.

Results

Descriptive statistics

Table 2. shows the descriptive statistics of all the six traits studied. A high but normal coefficient of variation (CV) was observed for HBW whereas the morphometric traits typically had lower CVs.

Table 2. Descriptive statistics of Phenotypic measurements

Traits	Mean	Min	Max	CV
Weight	106	28	261	33.79
Total length	18.2	11.5	26.8	11.03
Standard length	14.4	3.2	19.4	10.68
Head size	4.49	2.9	6.6	12.08
Body Height	5.79	2.6	8.7	13.63
Width	2.55	1.5	3.9	13.11

The average CVs of all traits for each strain combination are displayed in Table 3. The purebreds, S2S2 and S3S3, with an average harvest bodyweight of 138 ± 33 g and 135 ± 31 g respectively, were larger than all the cross-bred strains. Figure-1, Supplement 2, shows the distribution of records for each trait within each strain group. The figures indicate that the distribution of observations for each trait are not perfect normally distributed, but had more extreme observations than could be expected, indicating that both shooters and stunted fish occurred. This may indicate social interaction and/or stressed/infected fish.

Table-3. Mean and CV for each strain groups among all traits.

Traits	S1S1		S2S2		S3S3		S1S2		S2S3		S1S3	
	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV	Mean	CV
HBW	126.5	21.0	138.1	23.6	135.7	22.8	130.6	20.9	131.6	23.6	129.1	19.8
TL	19.49	6.76	19.9	5.89	19.8	6.66	19.7	6.63	19.6	6.9	19.5	6.28
SL	15.33	6.82	15.7	7.07	15.6	6.70	15.8	6.36	15.5	7.0	15.3	6.47
HL	4.86	8.16	4.9	7.23	4.88	8.17	4.9	7.93	4.93	7.9	4.8	7.69
BH	6.36	8.34	6.5	8.60	6.38	9.41	6.3	8.90	6.35	9.5	6.3	8.64
BW	2.64	10.9	2.8	14.3	2.82	11.27	2.6	10.3	2.72	11.6	2.71	10.6

ADM and SFM models

Goodness of fit

Likelihood ratio tests (LRT) were used to decide whether the full models, SFM and ADM, had better fit than a simpler model, for each of the traits. The results of LRT for the goodness of fit is shown in table 4. For all models with Sire, Dam, and Fullsib effects, the Dam component was negligible for all traits. Therefore, the SF model was considered as a best-fit model for most traits. The fullsib effect was significant for all traits except for SL and HL but most pronounced for TL, HBW and BWD. So, a model with only the sire effect included was preferred for SL and HL.

Table 4: Log likelihood values for various models tested and the outcome of Likelihood ratio tests (LRT) among them. The * is used to denote the model with the significantly best fit within the hierarchy of models.

<i>Models</i>	<i>HBW</i>	<i>TL</i>	<i>SL</i>	<i>BH</i>	<i>BW</i>	<i>HL</i>
S	-3734.87	-360.375	-159.913*	382.915	1044.44	684.269*
SM	-3734.87	-360.375	-159.913	382.915	1044.44	684.269
SF	-3731.79**	-354.06***	-159.09	384.814*	1048.69**	685.55
SFM	-3731.79	-354.064	-159.09	384.814	1048.69	685.55
A	-3817	-365.311	-170.532	368.46	1040.47	667.998
AD	-3816.86	-365.143	-169.116*	369.17	1044.23	668.482
AM	-3813.8**	-359.047***	-169.592	370.421*	1044.79	669.496*
ADM	-3813.72	-358.808	-168.165	370.834	1048.16***	669.931

*, ** and ***: The threshold values for Type-1 errors of *:0.05, **: 0.01, and ***:0.001 were: 1 d. f., 2.71, 5.42, and 9.55 respectively; for 2 d. f., 4.24, 7.29, and 11.77, respectively.

Variance components analysis

Table-1 in Supplements 1 shows all variances for all traits with all SFM and ADM models.

The estimates of different variance components, expressed as a proportion to the total phenotypic variance, are shown graphically in Figure 2. The simple model resulted in greater additive variance and heritability for all traits. By adding the dominance and maternal effects in the ADM models, the additive genetic variance in most cases decreased. And indeed, for HL, the additive variance was zero or around zero in both the full ADM and SFM models. No maternal ratios were estimated because of the negative estimates of maternal variance while using the SFM models.

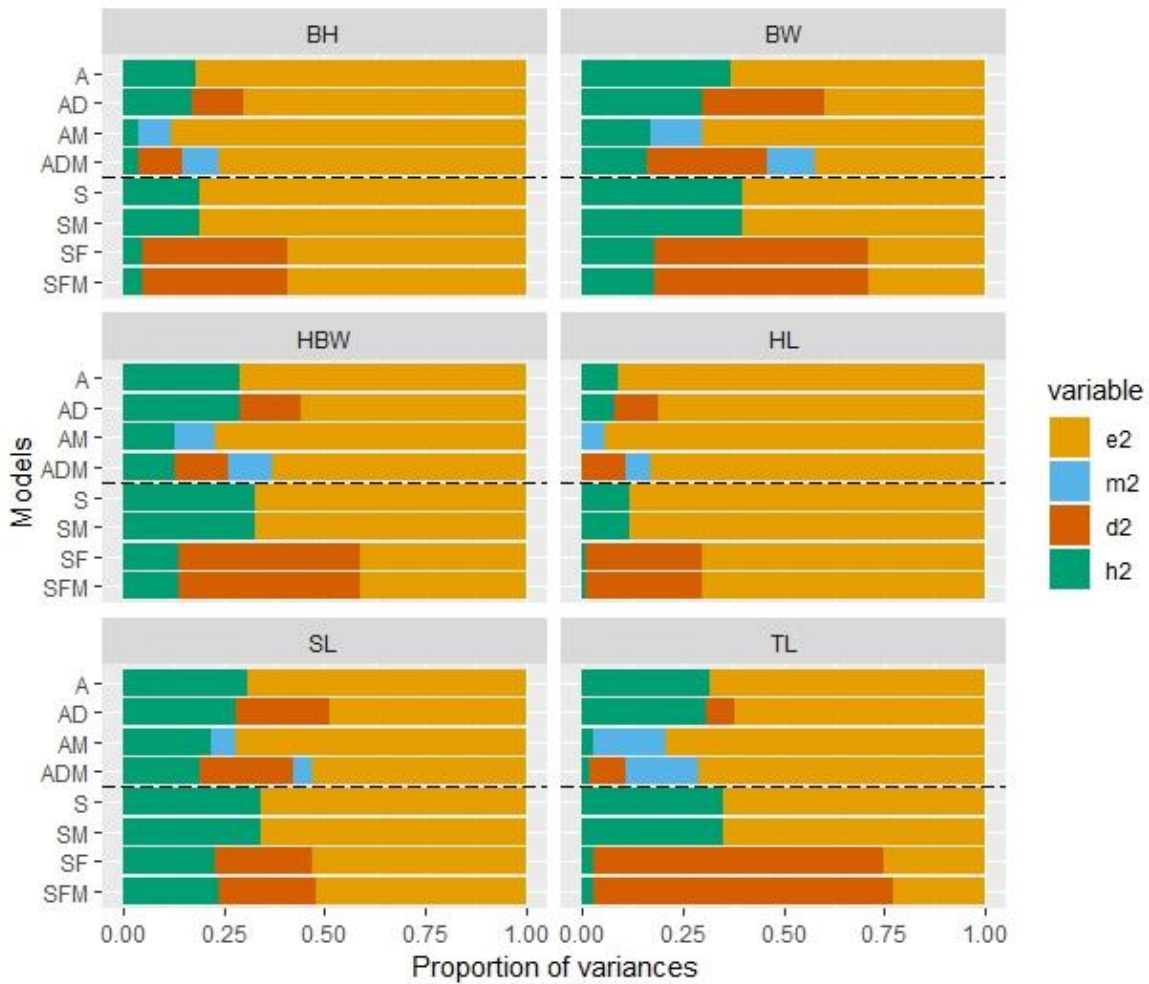


Figure 2: Segregation of the phenotypic variance of all traits into additive genetic (h^2), dominance (d^2), maternal (m^2), and error component (e^2) is shown in this figure. All nested models within ADM and SFM models are shown in the vertical axis of the plot.

Genetic parameters

The proportions of the additive variance, dominance variance and maternal variance are displayed in figure-2 and in table-1, Supplement1. The heritability, dominance ratio and the maternal ratio of the best fit models, SF and ADM models, are shown in Table 6.

The maternal ratio was not available for the SF model for any of the traits because of the negative estimates of maternal variance. However, the maternal effect was significant in all ADM models for all traits except SL. The dominance ratio was greater in a portion of total phenotype variance in all SF models compared to all ADM models. For BW, the dominance ratio was estimated as 0.30

± 0.13 using the ADM model, whereas it was 0.53 ± 0.22 when using the SF model. The heritability, h^2 , for all traits relied largely on the model fitted and had poor precision. The heritability for HBW, SL, and BW was moderate or greater than zero for the best fit models, as shown in Table 6. But the heritability of HL, TL and BH were not significantly greater than zero, which means very low estimates of heritability for those traits.

Table 6. Heritability, dominance ratio, and maternal Ratio \pm SE for the best fit models.

	Traits	Heritability	Dominance Ratio	Maternal Ratio
SF MODEL	HBW	0.14 \pm 0.11	0.45 \pm 0.21	-
	TL	0.03 \pm 0.11	0.72 \pm 0.29	-
	SL	0.34 \pm 0.07	-	-
	HL	0.12 \pm 0.08	-	-
	BH	0.05 \pm 0.09	0.36 \pm 0.20	-
	BW	0.18 \pm 0.11	0.53 \pm 0.22	-
ADM MODEL	HBW	0.13 \pm 0.09	-	0.10 \pm 0.05
	TL	0.03 \pm 0.11	-	0.18 \pm 0.06
	SL	0.28 \pm 0.06	0.23 \pm 0.14	-
	HL	0	-	0.06 \pm 0.03
	BH	0.04 \pm 0.08	-	0.08 \pm 0.05
	BW	0.16 \pm 0.10	0.30 \pm 0.13	0.12 \pm 0.06

SRH Model

In the SRH model, the additive genetic effects, general reciprocal strain effects and heterosis effects of each strain were measured as a fixed effect. A Wald-f test was performed to evaluate the fixed effects. Table-6 shows estimates of the additive genetic, reciprocal, and heterosis effects as obtained with the SRH model for the six traits investigated. The additive strain effects are given relative to the S3 strain, as this strain is put to zero by the ASReml program, whereas reciprocal and heterosis effects are shown as absolute values. The estimates of reciprocal effects and heterosis

effects were not significant between strains and strain combinations. All the crosses including the S1 strain gave negative estimates of the heterosis effect, whereas the crossbred S2S3 strain showed positive heterosis effect for HBW, BH, and HL. All heterosis effects were negative for TL, SL, and BW among all crossbred strains.

Table 6. Estimates of additive genetic, reciprocal and total heterosis effects of all phenotypic measurements among the crosses of three Mozambique Tilapia strains, (S1= Bons Sinais and Govuro, S2 = Marrangua, S3= Sotiva and Catuane)

EFFECTS		TRAITS				
RECIPROCAL	HBW	TL	SL	BH	HL	BW
S3	0.00	0.000	0.00	0.00	0.00	0.00
S2	2.32	-0.051	0.06	0.05	0.001	0.01
S1	2.01	0.027	-0.05	0.05	0.01	0.02
HETEROSIS						
S1S3×S3S1	-2.5	-0.045	-0.06	-0.04	0.04	-0.05
S2S3×S3S2	0.99	-0.041	-0.02	0.05	0.05	-0.02
S1S2×S2S1	-3.4	-0.132	-0.10	-0.04	0.01	-0.06
ADDITIVE GENETIC						
S3	0.00	0.000	0.00	0.00	0.00	0.00
S2	-7.38	-0.001	-0.08	-0.13	0.02	-0.10**
S1	-9.87***	-0.455***	-0.39***	-0.05	-0.07	-0.16***

Significance levels for test of estimates being different from zero: ***P<0.001, **P<0.01 and *P<0.05

Discussion

Additive and non-additive genetic strain effects

Among all the strain combinations, S2S3×S3S2 showed positive heterosis effect for HBW, thus the crossing between S2 and S3 pure breeds has the best future prospects in the breeding program. The other strain combinations had negative heterosis effects. None of the heterosis effects estimated for the different crossbred strains was statistically significant ($p < 0.05$) in this study. In contrast to this, Workagegn (2019) observed significant heterosis while crossing three different strains of *O. niloticus* in Ethiopia, but no significant additive genetic or reciprocal effects were observed among strain groups. In common carp, heterosis has commonly been observed among different cross-bred strains, which largely varied by genetic factors, genotype-environment interaction and age of the fish (Wohlfarth, 1993). Similarly, the heterosis effect of *O. mossambicus* might also be related to the age of the fish.

The additive genetic strain effects that are presented in this study are important because it represents the performance of a specific strain group, for each of the traits investigated. The results indicate that the S1 strain group had inferior performance compared to other strain groups in most of the traits. Also, little reciprocal effects were shown among strain groups. The reciprocal effects are quite common in many aquatic species and the effect can be confounded with fullsib or common environmental effect stemming from the mouth breeding nature of fish (Thanh et al., 2010). Bentsen et al. (1998) have observed significant additive genetic and reciprocal cross effects in different strains of *O. niloticus*, which were derived from a diallel cross of eight different strains of Nile Tilapia. The presence of some significant reciprocal effects conceptualises the theory that progenies of some dam strain groups could outperform the offspring of the sire strain groups. However, there was no strong evidence in our study to support the presence of reciprocal effect in *O. mossambicus* for the growth traits.

Heritability

In this study, heritability varied quite a lot for the traits measured. Heritability for HBW was about 0.3 ± 0.1 using the two simplest models (Table-1, Supplement 1), and this estimate is in accordance with other results that are reported earlier in various Tilapia species by other authors, where the heritabilities ranged from 0 to 0.42, as shown in table-1 in supplementary 6. These heritability estimates differ from the estimates obtained for Nile Tilapia for harvest body weight by (Joshi et

al., 2018) where the estimates were very close to zero for harvest body weight, with low standard error. The main reason for this was that with their design and models, much of the additive genetic variation were transformed into mostly epistatic effects. However, all the estimates of the heritability in our results were associated with large standard errors.

Non-additive genetic variance components

The components of dominance and maternal effects were estimated by using different models in this study, where the total phenotypic variance either had a large portion defined by dominance and a small or no portion described by the maternal effects. The variance components estimated by the SFM and ADM models which applies three basic components to explain the causal variation (Joshi et al., 2018), but the output of variances from both the models differed more in our study because the dam component was not estimated in any of the SFM models.

Maternal effects stem from genetic or non-genetic factors related to the dam, such as differences in egg size and egg quality caused by genetic or environmental differences during the sexual maturation (Cruz et al., 1997). As mentioned by Joshi (2018), the conventional experimental design causes confounding of the different variances but changing the mating system to e.g. a partial factorial design, can better separate the additive, maternal and non-additive genetic variances. However, even by using a full factorial design, he was not able to separate the variance components in an orthogonal manner in his analysis, even when applying a SNP-based genomic relationship matrix.

Best model fitted

The assumptions assumed when constructing the dominance relationship matrices ignore the fact that inbreeding decreases with the heterosis level, but if the inbreeding level is very low no serious bias will occur for the estimated dominance effects. The effect is expected to have a mean zero and symmetrical distribution. But Joshi (2018) points out that this is not completely true when directional dominance is present. No inbreeding was assumed in our study as the level of inbreeding is unknown in the wild sub-populations sampled. However, in general, compared to a pedigree-based BLUP method, genomic prediction methods with use of high density markers are the better option when trying to define these variances (Joshi, 2018).

Population subdivision

As illustrated by the heatmap, the three strains seem to be three clear different genetic groups. In addition, the S1 strain seems to be related to Nile Tilapia to some degree (Fig.1). However, the heatmap used in this study is based on markers having low call rate, especially among the groups, which thus will give a low level of precision when we try to deduce the population substructure. But the isolation of the fish sub-populations in the regions sampled give reasons to assume that subdivision of the population might have led to genetic drift or bottleneck effects, as the sudden occurrence of flood in the area could lead to genetic differences and homogenous isolated populations (Crispo & Chapman, 2009). On the other hand, relative levels of genetic diversity that was observed among subpopulations might also be caused by the habitats and soil types of the regions. According to Falconer and Mackay (1996), the gene frequency or genetic variation between subpopulations will influence the level of heterosis in crossbred populations generated from two specific lines or strains. In our study, we have presumed that S1, S2 and S3 strain groups had different genetic properties based on the available genotype and phenotype data.

Conclusion

The partial factorial mating design that was used in this breeding program for *O. mossambicus* allowed the separation of variance components pertaining to additive and non-additive genetic effects, and substantial amounts of dominance and maternal components were found. A major portion of the non-additive variation was defined by the dominance in this pedigree-based analysis. However, the standard error of the estimate of heritability, dominance and maternal components were large, which illustrated the fact that a large number of observations is required to obtain estimates with sufficient accuracy.

Potentially, genomic prediction methods based on high-quality markers will separate the additive genetic and non-additive genetic variance with more accuracy. Therefore, further approaches should be undertaken by using the genomic information based on high quality genotyping data in order to quantify the accurate non-additive sources of phenotypic variation for the desired trait.

Differences among strain groups in terms of additive genetic strain effects were also found. Based on the observed heterosis effects, the combination of the two strains S2 and S3 is most suitable for further genetic improvement based on their additive genetic levels and positive heterosis effect.

References

- Alda Ma. Salia , P. J. (2008). *ECONOMIC ANALYSIS OF SMALL-SCALE TILAPIA AQUACULTURE IN MOZAMBIQUE*. UNU-Fisheries Training Program: United Nations University.
- Arjona, G. T. (2018). *The domestication of O. mossambicus in a breeding program in Mozambique*: Norwegian University of Life Sciences.
- Bentsen, H. & Gjerde, B. (1994). *Design of fish breeding programs*, vol. 19.
- Bentsen, H. B., Eknath, A. E., Palada-de Vera, M. S., Danting, J. C., Bolivar, H. L., Reyes, R. A., Dionisio, E. E., Longalong, F. M., Circa, A. V., Tayamen, M. M., et al. (1998). Genetic improvement of farmed tilapias: growth performance in a complete diallel cross experiment with eight strains of *Oreochromis niloticus* ICLARM contribution No. 1074.1. *Aquaculture*, 160 (1): 145-173. doi: [https://doi.org/10.1016/S0044-8486\(97\)00230-5](https://doi.org/10.1016/S0044-8486(97)00230-5).
- Berg, P. & Henryon, M. (1998). *A comparison of mating designs for inference on genetic parameters in fish*. WCGALP.
- Butler, D., Cullis, B., Gilmour, A., Gogel, B. J. T. S. o. Q., Department of Primary Industries & Fisheries: Brisbane, Q. (2009). ASReml-R reference manual (version 3).
- Crispo, E. & Chapman, L. J. (2009). Temporal Variation in Population Genetic Structure of a Riverine African Cichlid Fish. *Journal of Heredity*, 101 (1): 97-106. doi: 10.1093/jhered/esp078 %J Journal of Heredity.
- Cruz, P., Ibarra, A. M. J. J. o. E. M. B. & Ecology. (1997). Larval growth and survival of two catarina scallop (*Argopecten circularis*, Sowerby, 1835) populations and their reciprocal crosses. 212 (1): 95-110.
- Cyrus, D., Vivier, L. & MacKay, C. J. I. R., Coastal Research Unit of Zululand. (2005). Lake St Lucia fish survey-December 2004. 118: 1-27.
- Dodds, K. G., McEwan, J. C., Brauning, R., Anderson, R. M., van Stijn, T. C., Kristjánsson, T. & Clarke, S. M. J. B. g. (2015). Construction of relatedness matrices using genotyping-by-sequencing data. 16 (1): 1047.
- Dupont-Nivet, M., Vandeputte, M., Haffray, P. & Chevassus, B. J. A. (2006). Effect of different mating designs on inbreeding, genetic variance and response to selection when applying individual selection in fish breeding programs. 252 (2-4): 161-170.
- Eknath, A. E., Tayamen, M. M., Palada-de Vera, M. S., Danting, J. C., Reyes, R. A., Dionisio, E. E., Capili, J. B., Bolivar, H. L., Abella, T. A., Circa, A. V., et al. (1993). Genetic improvement of farmed tilapias: the growth performance of eight strains of *Oreochromis niloticus* tested in different farm environments**ICLARM Contribution No. 709. In Gall, G. A. E. & Chen, H. (eds) *Genetics in Aquaculture*, pp. 171-188. Amsterdam: Elsevier.
- Falconer, D. & Mackay, T. J. U. L. G. (1996). *Introduction to quantitative genetics*. Essex.
- FAO. (2018). *The State of World Fisheries and Aquaculture 2018 - Meeting the sustainable development goals*. THE STATE OF THE WORLD. Rome: the Food and Agriculture Organization of the United Nations.
- Fernández, J., Toro, M. Á., Sonesson, A. K. & Villanueva, B. (2014). Optimizing the creation of base populations for aquaculture breeding programs using phenotypic and genomic data and its consequences on genetic progress. *Frontiers in genetics*, 5: 414-414. doi: 10.3389/fgene.2014.00414.
- Fiess, J. C., Kunkel-Patterson, A., Mathias, L., Riley, L. G., Yancey, P. H., Hirano, T. & Grau, E. G. (2007). Effects of environmental salinity and temperature on osmoregulatory ability, organic osmolytes, and plasma hormone profiles in the Mozambique tilapia (*Oreochromis mossambicus*). *Comparative Biochemistry and Physiology Part A: Molecular & Integrative Physiology*, 146 (2): 252-264. doi: <https://doi.org/10.1016/j.cbpa.2006.10.027>.

- Firmat, C., Alibert, P., Losseau, M., Baroiller, J.-F. & Schliewen, U. K. (2013). Successive Invasion-Mediated Interspecific Hybridizations and Population Structure in the Endangered Cichlid *Oreochromis mossambicus*. *PLOS ONE*, 8 (5): e63880. doi: 10.1371/journal.pone.0063880.
- Fisher, R. A. J. T. E. R. (1919). The causes of human variability. 10 (4): 213.
- Fjalestad, K. T. (2005). Breeding Strategies. In Gjedrem, T. (ed.) *Selection and Breeding Programs in Aquaculture*, pp. 145-158.
- García-Ruiz, A., Wiggans, G. R. & Ruiz-López, F. J. (2019). Pedigree verification and parentage assignment using genomic information in the Mexican Holstein population. *Journal of Dairy Science*, 102 (2): 1806-1810. doi: <https://doi.org/10.3168/jds.2018-15076>.
- Gjedrem, T., Gjølén, H. M. & Gjerde, B. (1991). Genetic origin of Norwegian farmed Atlantic salmon. *Aquaculture*, 98 (1): 41-50. doi: [https://doi.org/10.1016/0044-8486\(91\)90369-l](https://doi.org/10.1016/0044-8486(91)90369-l).
- Gjedrem, T. (2005). Breeding Plan. In Gjedrem, T. (ed.) *Selection and Breeding Programs in Aquaculture*: Springer.
- Gjedrem, T., Robinson, N. & Rye, M. (2012). The importance of selective breeding in aquaculture to meet future demands for animal protein: A review. *Aquaculture*, 350-353: 117-129. doi: <https://doi.org/10.1016/j.aquaculture.2012.04.008>.
- Gjedrem, T. & Robinson, N. J. A. S. (2014). Advances by selective breeding for aquatic species: a review. 5 (12): 1152.
- Gjerde, B. (2005). Design of Breeding Programs. In Gjedrem, T. (ed.) *Selection and Breeding Programs in Aquaculture*, pp. 173-195.
- Gjølén, H. M. & Bentsen, H. B. (1997). Past, present, and future of genetic improvement in salmon aquaculture. *ICES Journal of Marine Science*, 54 (6): 1009-1014. doi: 10.1016/S1054-3139(97)80005-7 %J ICES Journal of Marine Science.
- Holtmark, M., Sonesson, A. K., Gjerde, B. & Klemetsdal, G. J. A. (2006). Number of contributing subpopulations and mating design in the base population when establishing a selective breeding program for fish. 258 (1-4): 241-249.
- Joshi, R. (2018). Non-additive genetic effects in Nile tilapia.
- Joshi, R., Woolliams, J. A., Meuwissen, T. H. E. & Gjølén, H. M. (2018). Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity*, 120 (5): 452-462. doi: 10.1038/s41437-017-0046-x.
- Kennedy, B., Schaeffer, L. & Sorensen, D. J. J. o. D. S. (1988). Genetic properties of animal models. 71: 17-26.
- Khaw, H. L., Bovenhuis, H., Ponzoni, R. W., Rezk, M. A., Charo-Karisa, H. & Komen, H. J. A. (2009). Genetic analysis of Nile tilapia (*Oreochromis niloticus*) selection line reared in two input environments. 294 (1-2): 37-42.
- Kolstad, K. (2005). Estimating phenotypic and genetic parameters. In Gjedrem, T. (ed.) *Selection and Breeding Programs in Aquaculture*.
- Lhorente, J. P., Araneda, M., Neira, R. & Yáñez, J. M. (2019). Advances in genetic improvement for salmon and trout aquaculture: the Chilean situation and prospects. 11 (2): 340-353. doi: 10.1111/raq.12335.
- Maluwa, A. O. & Gjerde, B. (2006b). Genetic evaluation of four strains of *Oreochromis shiranus* for harvest body weight in a diallel cross. *Aquaculture*, 259 (1): 28-37. doi:<https://doi.org/10.1016/j.aquaculture.2006.06.003>.
- Makgahlela, M., Strandén, I., Nielsen, U., Sillanpää, M. & Mäntysaari, E. J. J. o. d. s. (2013). The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations. 96 (8): 5364-5375.

- Marjanovic, J., Mulder, H. A., Khaw, H. L. & Bijma, P. J. G. S. E. (2016). Genetic parameters for uniformity of harvest weight and body size traits in the GIFT strain of Nile tilapia. 48 (1): 41.
- Modadugu, V. G. & O. Acosta, B. (2004). *A review of global tilapia farming practices*, vol. 9.
- Muñoz, P. R., Resende, M. F., Gezan, S. A., Resende, M. D. V., de los Campos, G., Kirst, M., Huber, D. & Peter, G. F. J. G. (2014). Unraveling additive from nonadditive effects using genomic relationship matrices. 198 (4): 1759-1768.
- Nguyen, N. H., Ponzoni, R. W., Abu-Bakar, K. R., Hamzah, A., Khaw, H. L. & Yee, H. Y. J. A. (2010). Correlated response in fillet weight and yield to selection for increased harvest weight in genetically improved farmed tilapia (GIFT strain), *Oreochromis niloticus*. 305 (1-4): 1-5.
- Nielsen, H., Sonesson, A., Meuwissen, T. (2011). "Optimum contribution selection using traditional best linear unbiased prediction and genomic breeding values in aquaculture breeding schemes." *J Anim. Sci.* **89**: 630 - 638.
- Palucci, V., Schaeffer, L. R., Miglior, F. & Osborne, V. J. G. S. E. (2007). Non-additive genetic effects for fertility traits in Canadian Holstein cattle (Open Access publication). 39 (2): 181. doi: 10.1186/1297-9686-39-2-181.
- Pante, M. J. R., 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 (3): 383-392. doi: [https://doi.org/10.1016/S0044-8486\(01\)00825-0](https://doi.org/10.1016/S0044-8486(01)00825-0).
- Ponzoni, R. W., 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 (1): 203-210. doi: <https://doi.org/10.1016/j.aquaculture.2005.02.020>.
- Raidan, F. S. S., Porto-Neto, L. R., Li, Y., Lehnert, S. A., Vitezica, Z. G. & Reverter, A. (2018). Evaluation of nonadditive effects in yearling weight of tropical beef cattle1. *Journal of Animal Science*, 96 (10): 4028-4034. doi: 10.1093/jas/sky275 %J Journal of Animal Science.
- Robledo, D., Palaiokostas, C., Bargelloni, L., Martínez, P. & Houston, R. (2018). Applications of genotyping by sequencing in aquaculture breeding and genetics. 10 (3): 670-682. doi: 10.1111/raq.12193.
- Sørensen, A. C., Berg, P. & Woolliams, J. A. J. G. S. E. (2005). The advantage of factorial mating under selection is uncovered by deterministically predicted rates of inbreeding. 37 (1): 57. doi: 10.1186/1297-9686-37-1-57.
- Su, G., Christensen, O. F., Ostensen, T., Henryon, M. & Lund, M. S. (2012). Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLOS ONE*, 7 (9): e45293. doi: 10.1371/journal.pone.0045293.
- Thanh, N. M., Nguyen, N. H., Ponzoni, R. W., Vu, N. T., Barnes, A. C. & Mather, P. B. (2010). Estimates of strain additive and non-additive genetic effects for growth traits in a diallel cross of three strains of giant freshwater prawn (*Macrobrachium rosenbergii*) in Vietnam. *Aquaculture*, 299 (1): 30-36. doi: <https://doi.org/10.1016/j.aquaculture.2009.12.011>.
- Visscher, P. M. & Bruce Walsh, J. J. I. J. o. e. (2017). Commentary: Fisher 1918: the foundation of the genetics and analysis of complex traits. 48 (1): 10-12.
- Whitfield, A. K., Taylor, R. H., Fox, C. & Cyrus, D. P. (2006). Fishes and salinities in the St Lucia estuarine system—a review. *Reviews in Fish Biology and Fisheries*, 16 (1): 1. doi: 10.1007/s11160-006-0003-x.
- William, R. & Pollak, E. J. J. o. D. S. (1985). Theory of heterosis. 68 (9): 2411-2417.
- Wohlfarth, G. W. J. A. (1993). Heterosis for growth rate in common carp. 113 (1-2): 31-46.
- Wolak, M. E. (2012). nadv : an R package to create relatedness matrices for estimating non-additive genetic variances in animal models. 3 (5): 792-796. doi: 10.1111/j.2041-210X.2012.00213.x.
- Workagegn, K.B., 2019. Establishment of a base population for long-term genetic improvement of Nile tilapia in Ethiopia. NMBU, Philosophiae Doctor (PhD) Thesis 2019:8

Yamaguchi, Y., Breves, J. P., Haws, M. C., Lerner, D. T., Grau, E. G. & Seale, A. P. (2018). Acute salinity tolerance and the control of two prolactins and their receptors in the Nile tilapia (*Oreochromis niloticus*) and Mozambique tilapia (*O. mossambicus*): A comparative study. *General and Comparative Endocrinology*, 257: 168-176. doi: <https://doi.org/10.1016/j.ygcen.2017.06.018>.

Supplement 1

Table-1: The variance components estimation for the different models tested. Standard errors are shown in parenthesis.

Traits	Model	σ^2_s	σ^2_D	σ^2_F	σ^2_{add}	σ^2_{dom}	σ^2_{mat}	σ^2_e	σ^2_p	h^2	d^2	m^2
HBW	S	26.92 _(5.96)			107.70 _(23.87)			299.14 _(14.13)	326.07 _(13.97)	0.33 _(0.07)		
HBW	SD	26.92 _(5.96)	0		107.70 _(23.87)		NV	299.14 _(14.13)	326.07 _(13.97)	0.33 _(0.07)		NV
HBW	SF	12.32 _(8.86)		37.91 _(19.54)	49.26 _(35.38)	151.64 _(78.01)		290.16 _(14.05)	340.38 _(17.10)	0.14 _(0.11)	0.45 _(0.21)	
HBW	SDF	12.32 _(8.86)	0	37.91 _(19.54)	49.26 _(35.38)	151.65 _(78.01)	NV	290.16 _(14.05)	340.38 _(17.08)	0.14 _(0.11)	0.45 _(0.21)	NV
HBW	A				108.89 _(24.19)			257.01 _(20.98)	365.91 _(16.61)	0.29 _(0.06)		
HBW	AD				106.08 _(24.45)	24.79 _(46.76)		234.39 _(46.59)	365.97 _(16.61)	0.29 _(0.06)	0.15 _(0.12)	
HBW	AM				49.39 _(35.61)		32.92 _(19.71)	277.47 _(22.79)	365.52 _(16.58)	0.13 _(0.09)		0.10 _(0.05)
HBW	ADM				47.99 _(35.54)	19.08 _(46.54)	38.44 _(19.71)	260.07 _(47.81)	365.58 _(16.8)	0.13 _(0.09)	0.13 _(0.13)	0.11 _(0.05)
TL	S	0.058			0.23 _(0.052)			0.61 _(0.03)	0.66 _(0.03)	0.35 _(0.07)		
TL	SD	0.058	0		0.23 _(0.052)		NV	0.61 _(0.03)	0.66 _(0.03)	0.35 _(0.08)		NV
TL	SF	0.006		0.13	0.022 _(0.077)	0.53 _(0.03)		0.58 _(0.03)	0.74 _(0.067)	0.03 _(0.1)	0.72 _(0.29)	
TL	SDF	0.006	0	0.13	0.022 _(0.07)	0.53 _(0.19)	NV	0.58 _(0.03)	0.715 _(0.038)	0.03 _(0.11)	0.74 _(0.24)	NV
TL	A				0.23 _(0.053)			0.49 _(0.04)	0.73 _(0.034)	0.32 _(0.06)		
TL	AD				0.22 _(0.05)	0.055 _(0.09)		0.45 _(0.09)	0.73 _(0.03)	0.31 _(0.07)	0.07 _(0.13)	
TL	AM				0.03 _(0.07)		0.13 _(0.05)	0.57 _(0.05)	0.73 _(0.03)	0.03 _(0.1)		0.18 _(0.06)
TL	ADM				0.02 _(0.08)	0.06 _(0.09)	0.13 _(0.05)	0.51 _(0.09)	0.73 _(0.03)	0.02 _(0.11)	0.09 _(0.13)	0.18 _(0.06)
SL	S	0.039			0.15 _(0.035)			0.42 _(0.02)	0.46 _(0.02)	0.34 _(0.07)		
SL	SD	0.039	0		0.15 _(0.04)		NV	0.42 _(0.02)	0.46 _(0.02)	0.34 _(0.08)		NV
SL	SF	0.028		0.029	0.11 _(0.05)	0.11 _(0.1)		0.41 _(0.02)	0.47 _(0.02)	0.23 _(0.11)	0.24 _(0.21)	
SL	SDF	0.028	0	0.029	0.11 _(0.05)	0.11 _(0.1)	NV	0.41 _(0.02)	0.47 _(0.03)	0.24 _(0.11)	0.24 _(0.2)	NV
SL	A				0.16 _(0.04)			0.34 _(0.03)	0.51 _(0.02)	0.31 _(0.06)		
SL	AD				0.14 _(0.04)	0.12 _(0.07)		0.24 _(0.07)	0.51 _(0.02)	0.28 _(0.06)	0.23 _(0.14)	
SL	AM				0.11 _(0.05)		0.03 _(0.03)	0.37 _(0.03)	0.51 _(0.02)	0.22 _(0.09)		0.06 _(0.05)
SL	ADM				0.097 _(0.05)	0.12 _(0.07)	0.03 _(0.03)	0.26 _(0.07)	0.51 _(0.02)	0.19 _(0.09)	0.23 _(0.14)	0.05 _(0.05)
HL	S	0.003 _(0.001)			0.012 _(0.005)			0.095 _(0.004)	0.098 _(0.004)	0.12 _(0.05)		
HL	SD	0.003 _(0.001)	0		0.012 _(0.005)		NV	0.095 _(0.004)	0.098 _(0.004)	0.12 _(0.05)		NV
HL	SF	0.0002 _(0.002)		0.007 _(0.005)	0.0007 _(0.009)	0.029 _(0.02)		0.094 _(0.005)	0.10 _(0.005)	0.007 _(0.08)	0.29 _(0.19)	
HL	SDF	0.0002 _(0.002)	0	0.007 _(0.005)	0.0008 _(0.009)	0.029 _(0.02)	NV	0.094 _(0.005)	0.10 _(0.005)	0.007 _(0.08)	0.29 _(0.19)	NV
HL	A				0.009 _(0.005)			0.09 _(0.005)	0.104 _(0.005)	0.09 _(0.05)		
HL	AD				0.008 _(0.005)	0.01 _(0.01)		0.08 _(0.01)	0.104 _(0.005)	0.08 _(0.05)	0.11 _(0.13)	
HL	AM				0		0.006 _(0.003)	0.09 _(0.005)	0.10 _(0.005)	0		0.06 _(0.03)
HL	ADM				0	0.01 _(0.013)	0.006 _(0.003)	0.087 _(0.01)	0.104 _(0.005)	0	0.11 _(0.12)	0.06 _(0.03)
BH	S	0.009 _(0.002)			0.034 _(0.009)			0.161 _(0.007)	0.17 _(0.007)	0.19 _(0.06)		
BH	SD	0.009 _(0.002)	0		0.034 _(0.009)		NV	0.162 _(0.007)	0.17 _(0.007)	0.19 _(0.06)		NV
BH	SF	0.002 _(0.004)		0.016 _(0.009)	0.009 _(0.01)	0.064 _(0.038)		0.158 _(0.007)	0.176 _(0.009)	0.05 _(0.09)	0.36 _(0.2)	
BH	SDF	0.002 _(0.004)	0	0.016 _(0.009)	0.009 _(0.01)	0.064 _(0.038)	NV	0.158 _(0.007)	0.18 _(0.009)	0.05 _(0.09)	0.36 _(0.2)	NV
BH	A				0.033 _(0.009)			0.15 _(0.01)	0.18 _(0.008)	0.18 _(0.05)		
BH	AD				0.031 _(0.01)	0.024 _(0.024)		0.13 _(0.02)	0.18 _(0.008)	0.17 _(0.05)	0.13 _(0.13)	
BH	AM				0.008 _(0.016)		0.016 _(0.009)	0.15 _(0.01)	0.18 _(0.008)	0.04 _(0.08)		0.08 _(0.05)

<i>BH</i>	ADM			0.008 _(0.016)	0.02 _(0.02)	0.016 _(0.009)	0.14 _(0.024)	0.18 _(0.008)	0.04 _(0.08)	0.11 _(0.13)	0.09 _(0.05)
<i>BW</i>	S	0.005 _(0.001)		0.02 _(0.004)			0.045 _(0.002)	0.05 _(0.002)	0.4 _(0.08)		
<i>BW</i>	SD	0.005 _(0.001)	0	0.02 _(0.004)		NV	0.045 _(0.002)	0.05 _(0.002)	0.4 _(0.08)		NV
<i>BW</i>	SF	0.002 _(0.001)	0.007 _(0.003)	0.009 _(0.005)	0.028 _(0.013)		0.043 _(0.002)	0.05 _(0.003)	0.18 _(0.11)	0.53 _(0.22)	
<i>BW</i>	SDF	0.002 _(0.001)	0	0.007 _(0.003)	0.009 _(0.005)	0.028 _(0.013)	NV	0.043 _(0.002)	0.053 _(0.003)	0.18 _(0.11)	0.53 _(0.22)
<i>BW</i>	A			0.02 _(0.004)			0.03 _(0.003)	0.05 _(0.003)	0.37 _(0.06)		
<i>BW</i>	AD			0.02 _(0.004)	0.02 _(0.008)		0.018 _(0.007)	0.056 _(0.003)	0.3 _(0.07)	0.3 _(0.13)	
<i>BW</i>	AM			0.0095 _(0.006)		0.007 _(0.003)	0.04 _(0.004)	0.056 _(0.003)	0.17 _(0.10)		0.13 _(0.06)
<i>BW</i>	ADM			0.009 _(0.006)	0.02 _(0.008)	0.007 _(0.003)	0.06 _(0.003)	0.06 _(0.003)	0.16 _(0.10)	0.30 _(0.13)	0.12 _(0.06)

Supplement 2

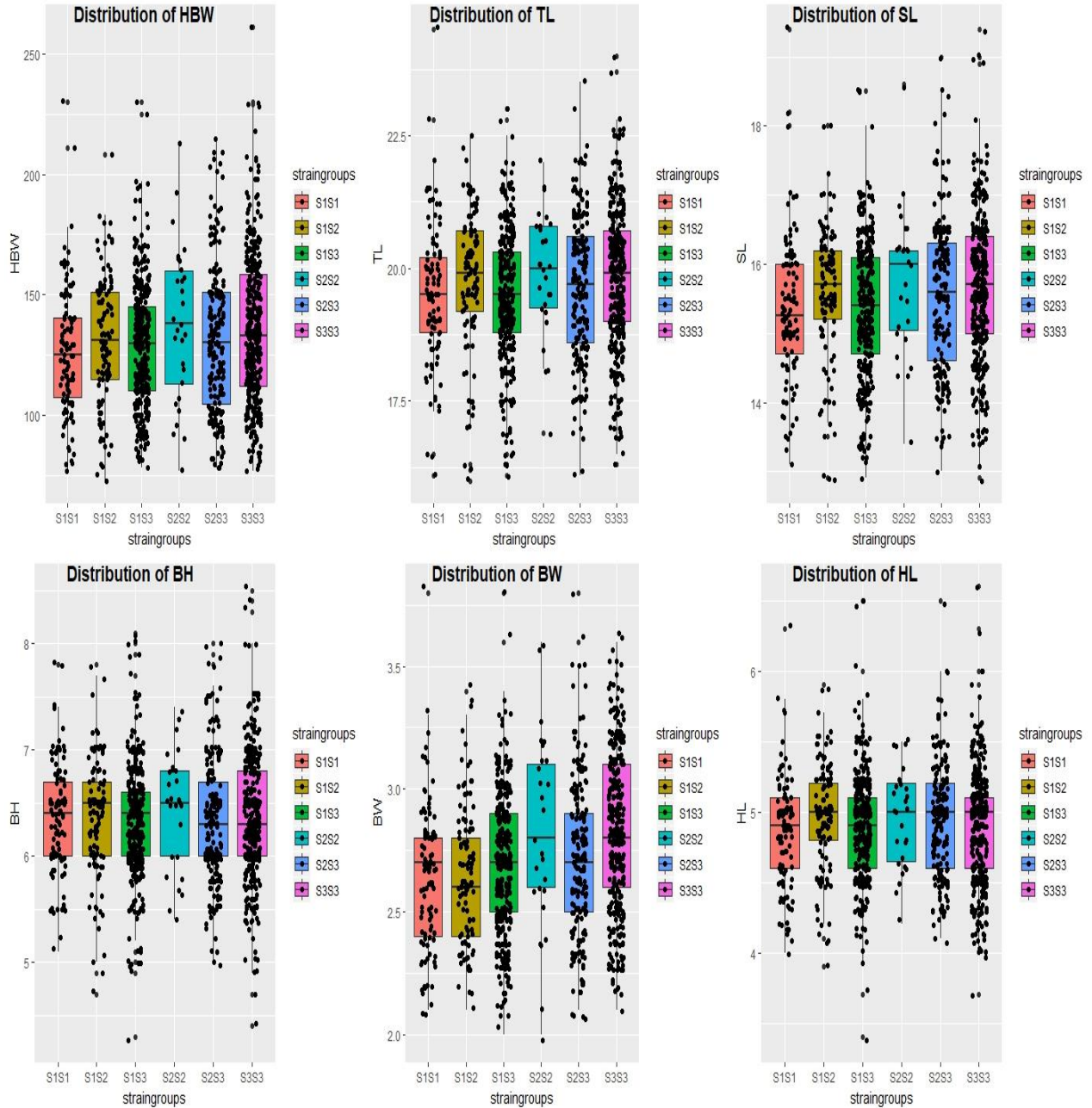


Figure-1. The boxplots of the distribution of different phenotypes for each strain groups. The horizontal bar indicates the median value of each strain group and each colored rectangular represents the 25th to 75th percentile of the values of each strain group. And each dot corresponds to the phenotype of an individual strain, which is plotted on the y-axis.

Supplement 3

A separate list of hypothetical sires and dams with their respective relatedness values from G-Matrix were created batch-by-batch for each offspring. Hypothetical sires and dams were targeted to match as a couple enlisted in the original mating list of a specific batch. If they were matched with an original mating list along with their batch information, then they were considered as parents for that offspring.

1st step: To match with the mating list, 1st maximum relatedness values of the hypothetical parent were considered. Hypothetical sire having 1st maximum relatedness value and hypothetical dam having 1st maximum relatedness value for an offspring were considered as a couple for a specific batch. And, the couple was checked with the original mating list for a specific batch, if that sire-dam combination existed. Later, 2nd, 3rd, and 4th maximum relatedness values of parents were considered to match with the mating list of a specific batch.

2nd Step: Offspring without any parentage assignment were separated. Now, hypothetical sire's nearby relatedness value with dam was considered. For example, Hypothetical sire having 1st maximum relatedness value and hypothetical dam having 2nd maximum relatedness value were considered as a partner for a specific batch. If it was not matched with the mating list, then 2nd maximum, 3rd maximum, and 4th maximum relatedness values were considered to match with the mating list like the previous step.

After making several sire and dam combinations mentioned in the previous steps, the rest of the unmatched offspring were separated. For the rest unmatched offspring, a separate manual approach was being used to assign parents among them.

Figure-1. Description of the parentage assignment by using estimates of G-Matrix

Supplement 4

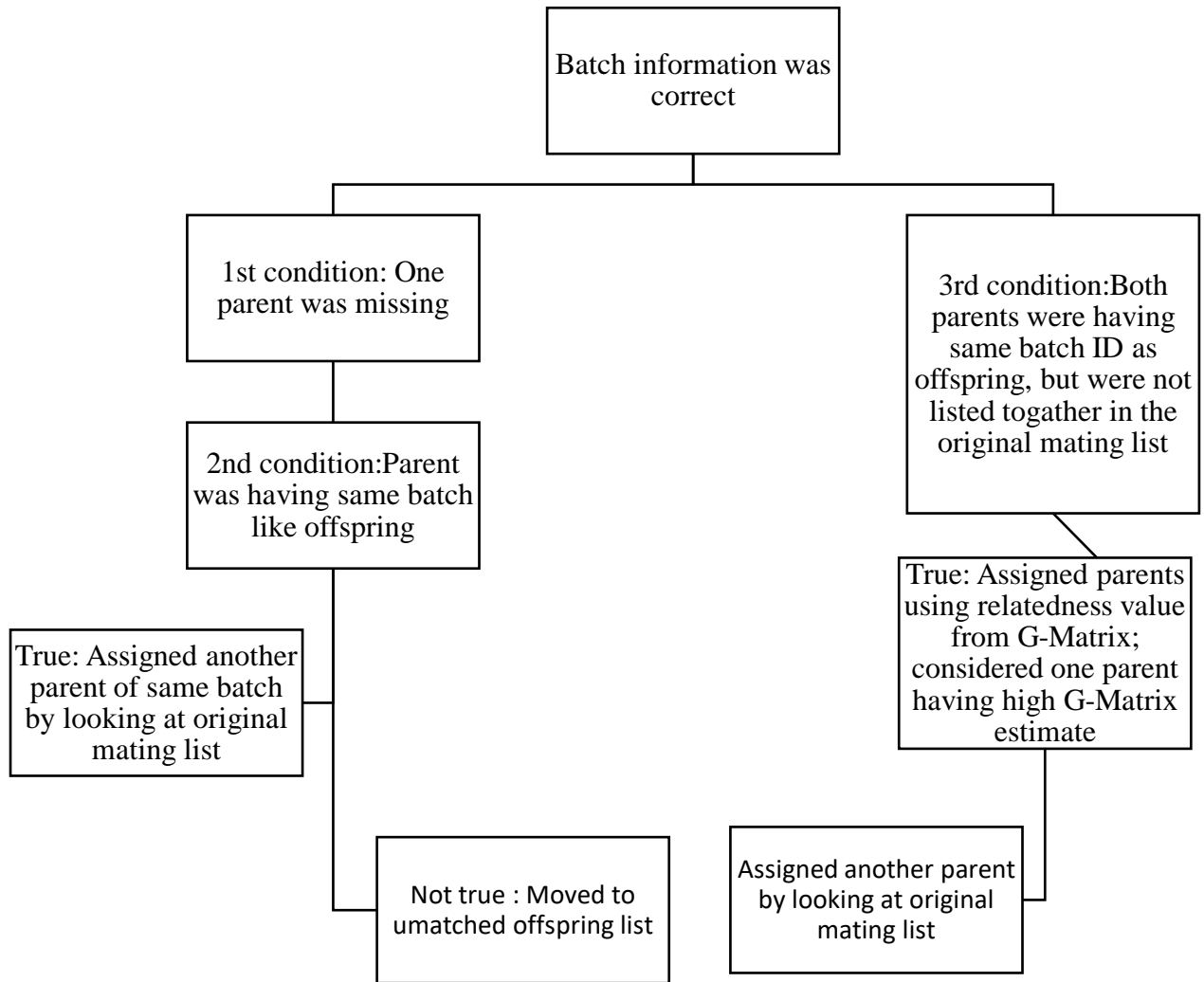


Figure-1. The manual steps for parentage assignment when the batch information is correct.

Supplement 5

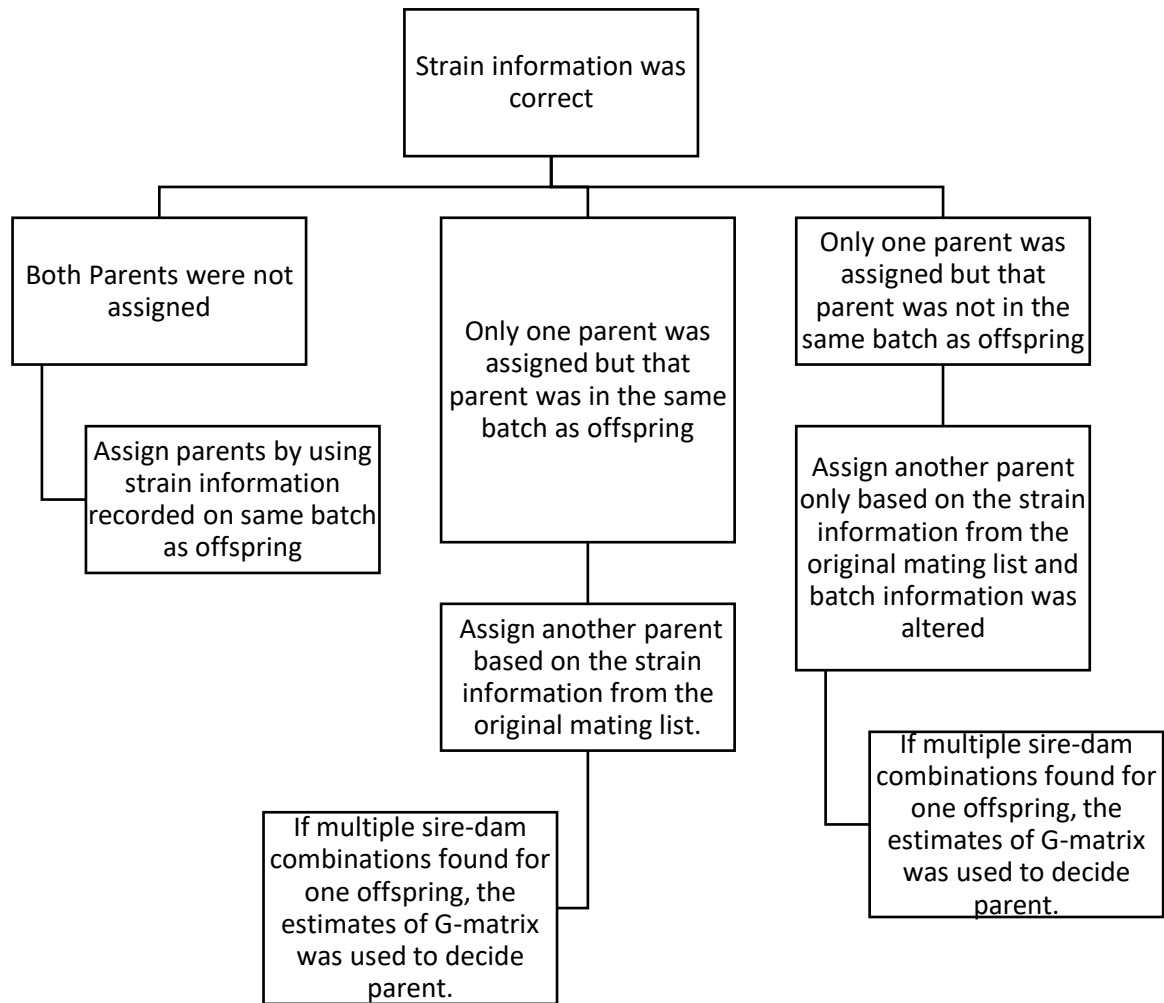


Figure-1. The manual steps for parentage assignment when the strain information is correct.

Supplement 6

Table-1: Literature review was performed for the heritability of Body weight related traits in Nile Tilapia.

Heritability	Trait	Species	Authors
0.23(0.08)	Daily weight gain	Nile Tilapia	(García-Ruiz et al., 2019)
0.41(0.22)	Bodyweight at Slaughter	Nile tilapia	(García-Ruiz et al., 2019)
0.25 (0.04)	Harvest body weight	Nile Tilapia (GIFT)	(Marjanovic et al., 2016)
0 or 0.06 (0.03)	Harvest Bodyweight	Nile Tilapia	(Joshi et al., 2018)
0.15 to 0.41 (0.04)	Harvest Weight	Nile Tilapia	(Khaw et al., 2009)
0.31(0.11)	Bodyweight	Nile Tilapia (GIFT)	(Nguyen et al., 2010)
0.34 (0.069)	Live Weight	Nile Tilapia	(Ponzoni et al., 2005)

Supplement 7

Table-1: Coefficients in the first generation (F1) derived from a diallel cross, for additive genetic (A_i), for reciprocal (R_i), and for total heterosis strain cross (h_j) effects

Strain cross	A_i			R_i			h_j		
	S1	S2	S3	S1	S2	S3	S1S2×S2S1/ S2S1×S1S2	S2S3×S3S2/ S3S2×S2S3	S1S3×S3S1/ S3S1×S1S3
S1×S1	1	0	0	1	0	0	0	0	0
S1×S2	0.5	0.5	0	1	0	0	1	0	0
S1×S3	0.5	0	0.5	1	0	0	0	0	1
S2×S2	0	1	0	0	1	0	0	0	0
S2×S1	0.5	0.5	0	0	1	0	1	0	0
S2×S3	0	0.5	0.5	0	1	0	0	1	0
S3×S3	0	0	1	0	0	1	0	0	0
S3×S1	0.5	0	0.5	0	0	1	0	0	1
S3×S2	0	0.5	0.5	0	0	1	0	1	0

Supplement 8

S8- 1. R-code: Recoding strain and maternal effects

#Loading the phenotypic file.

```
alltotal<-read.csv(file.choose(),header = T)
```

```
refinedall<-alltotal[,c(1:6,27:30)]
```

Changing the Sire strain code

```
refinedall$Sire_str<-ifelse(refinedall$Sire_str=="A","S3",
  ifelse(refinedall$Sire_str=="C","S3",
    ifelse(refinedall$Sire_str=="D","S2",
      ifelse(refinedall$Sire_str=="B","S1",
        ifelse(refinedall$Sire_str=="E","S1",NA)
      )
    )
  )
))
```

##Changing the Dam Strain code

```
refinedall$Dam_str<-ifelse(refinedall$Dam_str=="A","S3",
  ifelse(refinedall$Dam_str=="C","S3",
    ifelse(refinedall$Dam_str=="D","S2",
      ifelse(refinedall$Dam_str=="B","S1",
        ifelse(refinedall$Dam_str=="E","S1",NA)
      )
    )
  )
))
```

#Making the unique Sire and Dam Strain combination code for three strains

```
refinedall$combination<-paste(refinedall$Sire_str,refinedall$Dam_str,sep = "")
table(refinedall$combination)
refinedall$combination<-ifelse(refinedall$combination=="S2S1","S1S2",
  ifelse(refinedall$combination=="S3S2","S2S3",
    ifelse(refinedall$combination=="S3S1","S1S3",
      ifelse(refinedall$combination=="S1S1","S1S1",
        ifelse(refinedall$combination=="S2S2","S2S2",
          ifelse(refinedall$combination=="S3S3","S3S3",
            ifelse(refinedall$combination=="S1S3","S1S3",
              ifelse(refinedall$combination=="S2S3","S2S3",
                ifelse(refinedall$combination=="S1S2","S1S2",
                  "
                    ,NA)))))))))
library(dplyr)
```

Setting up the code for Strain additive genetic effects

```
refinedall$S1S1<-refinedall$combination
refinedall$S1S1<-ifelse(refinedall$S1S1=="S1S1",1,
                        ifelse(refinedall$S1S1=="S1S2",0.5,
                              ifelse(refinedall$S1S1=="S1S3",0.5,0)))
table(refinedall$S1S1)
refinedall$S2S2<-refinedall$combination
refinedall$S2S2<-ifelse(refinedall$S2S2=="S2S2",1,
                        ifelse(refinedall$S2S2=="S2S3",0.5,
                              ifelse(refinedall$S2S2=="S1S2",0.5,0)))
table(refinedall$S2S2)
refinedall$S3S3<-refinedall$combination
refinedall$S3S3<-ifelse(refinedall$S3S3=="S3S3",1,
                        ifelse(refinedall$S3S3=="S2S3",0.5,
                              ifelse(refinedall$S3S3=="S1S3",0.5,0)))
table(refinedall$S3S3)
```

##Setting up the code for heterosis effects

```
refinedall$hS1S2<-refinedall$combination
refinedall$hS1S2<-ifelse(refinedall$hS1S2=="S1S2",1,0)
table(refinedall$hS1S2)
refinedall$hS2S3<-refinedall$combination
refinedall$hS2S3<-ifelse(refinedall$hS2S3=="S2S3",1,0)
table(refinedall$hS2S3)
refinedall$hS1S3<-refinedall$combination
refinedall$hS1S3<-ifelse(refinedall$hS1S3=="S1S3",1,0)
table(refinedall$hS1S3)
```

#Setting up the code for Reciprocal effects

```
refinedall$mS1<-refinedall$Dam_str
refinedall$mS1<-ifelse(refinedall$mS1=="S1",1,0)
table(refinedall$mS1)
refinedall$mS2<-refinedall$Dam_str
refinedall$mS2<-ifelse(refinedall$mS2=="S2",1,0)
table(refinedall$mS2)
refinedall$mS3<-refinedall$Dam_str
refinedall$mS3<-ifelse(refinedall$mS3=="S3",1,0)
table(refinedall$mS3)
colnames(refinedall)
refinedall<-refinedall[,-c(11,18)]
write.csv(refinedall,file = "pheno3strain.csv")
refinedall1<-refinedall[,-c(6)]
colnames(refinedall1)
refinedall1<-refinedall1[,c(1:5,10:19,6:9)]
```



```

library(data.table)
refinedall1<-setnames(refinedall1,"S1S1","aS1")
refinedall1<-setnames(refinedall1,"S2S2","aS2")
refinedall1<-setnames(refinedall1,"S3S3","aS3")
write.table(refinedall1,file = "pheno3strain.txt",sep = "\t")
summary(refinedall1)
table(refinedall1$aS1)
write.csv(refinedall1,file = "pheno3strain.csv")

```

S8-2 . Making the dominance relationship matrix for ASReml

```

library(nadiv)
pedigree<-read.delim(file.choose(),header = T) # loading the pedigree file
pednew<-prepPed(pedigree) # making the pedigree file suitable for the package
listD<-makeD(pednew, parallel = F)$listDinv # making the dominance matrix
write.table(listD,file = "domAsparse.giv",row.names = F)

```

S8-3. Command file for the SRH model in ASReml

```

Model for SRH
Id !P
Sire !A
Dam !A
Sirestr !A
Damstr !A
Batch !A 10 !DV*
Eggdate !A
Weight !DV*
combination !A
aS1
aS2
aS3
hS1S2
hS2S3
hS1S3
mS1
mS2
mS3
Tot_L !DV*
Stnd_L !DV*
Head !DV*
Heigth !DV*

```

Width !DV*

```
pedigree.txt !SKIP 1 !Make  
phenommodifiedV2.csv !SKIP 1 !CSV !AISING !MAXIT=1000
```

```
Weight ~ mu Batch aS3 aS2 aS1 hS1S3 hS2S3 hS1S2 mS3 mS2 mS1 !r Id  
#Tot_L ~ mu Batch aS1 aS2 aS3 hS1S2 hS2S3 hS1S3 mS1 mS2 mS3 !r Id  
#Stnd_L ~ mu Batch aS1 aS2 aS3 hS1S2 hS2S3 hS1S3 mS1 mS2 mS3 !r Id  
#Heigth ~ mu Batch aS1 aS2 aS3 hS1S2 hS2S3 hS1S3 mS1 mS2 mS3 !r Id  
#Head ~ mu Batch aS1 aS2 aS3 hS1S2 hS2S3 hS1S3 mS1 mS2 mS3 !r Id  
#Width ~ mu Batch aS1 aS2 aS3 hS1S2 hS2S3 hS1S3 mS1 mS2 mS3 !r Id
```

S8-4, Command File for the SFM models in ASReml

##Similar methods for all traits

for All SFM models

```
Id !P  
Sire !A  
Dam !A  
Sirestr !A  
Damstr !A  
Batch !A 10 !DV*  
Eggdate !A  
Weight !DV*  
combination !A  
aS1  
aS2  
aS3  
hS1S2  
hS2S3  
hS1S3  
mS1  
mS2  
mS3  
Tot_L !DV*  
Stnd_L !DV*  
Head !DV*  
Heigth !DV*  
Width !DV*
```

```
pedigree.txt !SKIP 1 !Make  
phenommodifiedV2.csv !SKIP 1 !CSV !AISING !MAXIT=1000
```

```
# Weight ~ mu Batch !r Sire and(Dam) fac(Sire,Dam)  
# Weight ~ mu Batch !r Sire and(Dam) ide(Dam)
```

```
# Weight ~ mu Batch !r Sire and(Dam)
# Weight ~ mu Batch !r Sire and(Dam) ide(Dam) fac(Sire,Dam)
```

```
VPREDICT !DEFINE
```

```
F VarA 1 * 4
```

```
#F VarC 2 - 1
```

```
F VarC 3 - 1
```

```
F VarD 2 * 4
```

```
F VarP 1 +2+ 3 + 4
```

```
H h2 VarA VarP
```

```
H c2 VarC VarP
```

```
H d2 VarD VarP
```

S8-5, Command File for the ADM models in ASReml

##Similar methods for all traits

Model without female with dominance matrix

```
Id !P
```

```
Sire !p
```

```
Dam !A
```

```
Sirestr 3 !A
```

```
Damstr 3 !A
```

```
Batch 10 !A
```

```
Eggdate !A
```

```
Weight
```

```
combi !A
```

```
aS1
```

```
aS2
```

```
aS3
```

```
hS1S2
```

```
hS2S3
```

```
hS1S3
```

```
mS1
```

```
mS2
```

```
mS3
```

```
Tot_L !DV*
```

```
Stnd_L !DV*
```

```
Head !DV*
```

```
Heigth !DV*
```

```
Width !DV*
```

```
pedigree.txt !SKIP1 !make
```

```
domAsparse.giv !SKIP1
```

```
phenommodifiedV2.csv !SKIP1 !AISING !MAXIT=1000
```

```
#Weight ~ mu Batch !r Id
```

```

#Weight ~ mu Batch !r Id fac(Sire,Dam)
#Weight~ mu Batch !r Id giv(Id)
#Weight ~ mu Batch !r Id giv(Id) fac(Sire,Dam)
Weight ~ mu Batch !r Id

```

```

!PIN !DEFINE
#F VarA 1 # Subjected to change based on the model
F VarA 2
F VarC 1
#F VarC 2 #Subjected to change based on the model
F VarD 3
#F VarD 2 #Subjected to change based on the model
F VarP 1 + 2 + 3 + 4
H h2 VarA VarP
H c2 VarC VarP
H D2 VarD VarP

```

S8- 6, For the parentage assignment – R codes.

#loading the csv, genomic relationship matrix file

```

matrix<-read.csv(file.choose(),header = TRUE, sep = ",")
row.names(matrix)
colnames(matrix)
list<-data.frame(matrix$seqID)
list$seq<-1:nrow(list)
library(data.table)
list<-setnames(list,"matrix.seqID","seqID")

```

#loading the converter files. For the batch information

```

tilapiasample<-read.csv(file.choose(),header = TRUE)
tilapiasample<-tilapiasample[,c(2,3,6)]
combinedlist<-merge(list,tilapiasample,by="seqID",all=TRUE)

```

Seperating the matrix based on sire and dam

```

parentcode<-read.csv(file.choose(),header = TRUE)
parentcode<-merge(parentcode,sireanddamcode,by.x = "Parents",by.y = "Tag_ID",all.x = TRUE)
parentcode<-parentcode[!duplicated(parentcode$Parents),]

```

```

parentcode<-na.omit(parentcode)

library(Hmisc)
submatrix<-matrix[which(rownames(matrix)%in%parentcode$seqID),]
submatrix<-submatrix[,which(colnames(submatrix)%nin%parentcode$seqID)]
submatrix[1:4,1:4]
submatrix<-merge(submatrix,parentcode,by = "seqID")
submatrix[1:4,2236:2240]

library(dplyr)
siresubmatrix<-filter(submatrix,code=="s")
damsubmatrix<-filter(submatrix,code=="d")

##loading the maxn function for the maximum relatedness value
maxn<-function(n)function(x)order(x,decreasing = TRUE)[n]

#1st maximum relatedness value of sire results
sire1results<-apply(siresubmatrix[,c(2:2237)],2,maxn(1))
sire1results<-data.frame(sire1results)
sire1results$value<-apply(siresubmatrix[,c(2:2237)],2,function(x)x[maxn(1)(x)])
sire1results<-setnames(sire1results,"sire1results","seq"
)
sire1results$offspring<-rownames(sire1results)
colnames(sire1results)
siresubmatrix[1:4,2236:2240]
submatrixsirelist<-siresubmatrix[,c(2238)]
submatrixsirelist<-data.frame(submatrixsirelist)
submatrixsirelist$seq<-1:nrow(submatrixsirelist)
colnames(submatrixsirelist)

```

```

colnames(submatrixsirelist)
submatrixsirelist$value<-NULL
submatrixsirelist$offspring<-NULL
sire1results$submatrixsirelist<-NULL
combinesire1<-merge(submatrixsirelist,sire1results,by="seq")
#1st maximum relatedness value of dam results
damsubmatrix[1:4,2236:2240]
submatrixdamlist<-damsubmatrix[,c(2238)]
submatrixdamlist<-data.frame(submatrixdamlist)
submatrixdamlist$seq<-1:nrow(submatrixdamlist)
dam1results<-apply(damsubmatrix[,c(2:2237)],2,maxn(1))
dam1results<-data.frame(dam1results)
dam1results$value<-apply(damsubmatrix[,c(2:2237)],2,function(x)x[maxn(1)(x)])
dam1results<-setnames(dam1results,"dam1results","seq"
)
dam1results$offspring<-rownames(dam1results)
colnames(dam1results)
colnames(submatrixdamlist)
submatrixdamlist$value<-NULL
submatrixdamlist$offspring<-NULL
dam1results$submatrixdamlist<-NULL
combinedam1<-merge(submatrixdamlist,dam1results,by="seq")
combinedsiredam1<-merge(combinesire1,combinedam1,by="offspring")
#add the batch(dont need to repeat)
batch<-read.csv(file.choose(),header = TRUE)
batch<-merge(batch,combinedlist,by="Tag_ID",all.x = TRUE)
batch<-na.omit(batch)
#tagging with batch

```

```

combinedsiredam1<-merge(combinedsiredam1,batch,by.x="offspring",by.y = "seqID",all.x =
TRUE,all.y = FALSE)

combinedsiredam1<-na.omit(combinedsiredam1)

combinedsiredam1$testcode<-
paste(combinedsiredam1$submatrixsirelist,combinedsiredam1$submatrixdamlist,combinedsiredam1$Batch_ID,sep="")

####loading the main results (do not need to repeat)

mainresults<-read.csv(file.choose(),header = TRUE)

mainresults$maincode<paste(mainresults$SIRE,mainresults$DAM,mainresults$Batch_ID,sep="")
)

combinedsiredam1$results<-combinedsiredam1$testcode%in%mainresults$maincode
as.data.frame(table(combinedsiredam1$results))

matchedlist1<-dplyr::filter(combinedsiredam1,results=="TRUE")

rm(combineddam1,combinesire1,sire1results,dam1results,siresubmatrix,damsubmatrix)

#Writing the file with 1st maximum relatedness value of parent

write.csv(matchedlist1,file = "Gmatchwith1stmaxsire1stmaxdam.csv")

#refining the submatrix

submatrix<-submatrix[,which(colnames(submatrix)%nin%matchedlist1$offspring)]

#getting sire&damsubmatrix

library(dplyr)

siresubmatrix<-filter(submatrix,code=="s")

damsubmatrix<-filter(submatrix,code=="d")

#getting the sire1results (sire with first maximum relatedness value)

submatrix[1:4,445:448]

sire1results<-apply(siresubmatrix[,c(2:445)],2,maxn(1))

sire1results<-data.frame(sire1results)

sire1results$value<-apply(siresubmatrix[,c(2:445)],2,function(x)x[maxn(1)(x)])

library(data.table)

sire1results<-setnames(sire1results,"sire1results","seq"
)

```

```

sire1results$offspring<-rownames(sire1results)
colnames(sire1results)
siresubmatrix[1:4,445:448]
submatrixsirelist<-siresubmatrix[,c(446)]
submatrixsirelist<-data.frame(submatrixsirelist)
submatrixsirelist$seq<-1:nrow(submatrixsirelist)
colnames(submatrixsirelist)
submatrixsirelist$value<-NULL
submatrixsirelist$offspring<-NULL
sire1results$submatrixsirelist<-NULL
combinesire1<-merge(submatrixsirelist, sire1results,by="seq")
###getting the dam2 results (Dam with second maximum relatedness values)
#seprating the dam matrix
colnames(damsubmatrix)
damsubmatrix[1:4,444:448]
submatrixdamlist<-damsubmatrix[,c(446)]
submatrixdamlist<-data.frame(submatrixdamlist)
submatrixdamlist$seq<-1:nrow(submatrixdamlist)
#2nd dam results
dam2results<-apply(damsubmatrix[,c(2:445)],2,maxn(2))
dam2results<-data.frame(dam2results)
dam2results$value<-apply(damsubmatrix[,c(2:445)],2,function(x)x[maxn(2)(x)])
dam2results<-setnames(dam2results,"dam2results","seq")
dam2results$offspring<-rownames(dam2results)
colnames(dam2results)
colnames(submatrixdamlist)
submatrixdamlist$value<-NULL
submatrixdamlist$offspring<-NULL

```



```

dam2results$submatrixdamlist<-NULL

combinedam2<-merge(submatrixdamlist,dam2results,by="seq")

combinedsiredam2<-merge(combinesire1,combinedam2,by="offspring")

#tagging with batch

combinedsiredam2<-merge(combinedsiredam2,batch,by.x="offspring",by.y = "seqID",all.x =
TRUE,all.y = FALSE)

##matching the results with original mating list

combinedsiredam2$testcode<-
paste(combinedsiredam2$submatrixsirelist,combinedsiredam2$submatrixdamlist,combinedsiredam2$Batch_ID,sep="")

combinedsiredam2$results<-combinedsiredam2$testcode%in%mainresults$maincode
as.data.frame(table(combinedsiredam2$results))

matchedlist2<-dplyr::filter(combinedsiredam2,results=="TRUE")

# writing the output file with 1st and 2nd maximum relatedness value of parent

write.csv(matchedlist2,file = "Gmatch2with1stmaxsire2ndmaxdam.csv")

```




Norges miljø- og biovitenskapelige universitet
Noregs miljø- og biovitenskapelige universitet
Norwegian University of Life Sciences

Postboks 5003
NO-1432 Ås
Norway