Norges miljø- og
biovitenskapelige
universitet

Master's Thesis 2022 60 ECTS
Department of Animal and Aquacultural Sciences

Accuracy of Genomic Prediction using group recordings

Niran Adhikari
Animal Breeding and Genetics

SUPERVISOR

Professor Theodorus H.E. Meuwissen

Department of Animal and Aquacultural Science

Norwegian University of Life Sciences (NMBU)

Ås, Norway

ACKNOWLEDGEMENT

My sincere gratitude goes to Professor Theodorus Meuwissen, who accepted to supervise this project remotely. In my case, it helped me in many ways and unlocked the door to completing the master's degree. His guidance and support enabled me to stay on the research track. Further, his feedback assisted me in preparing the current form of the thesis. So I especially thank him. Also, I am thankful to my family members for their unending love and care.

CONTENTS

LIST OF ABBREVIATIONS

FCR: Feed Conversion Ratio

EBV: Estimated Breeding Value

TBV: True Breeding Value

BLUP: Best Linear Unbiased Prediction

SNP: Single nucleotide polymorphism

WGS: Whole Genome Sequence

Bayes: Bayesian

QTL: Quantitative Trait Loci

LD: Linkage Disequilibrium

GBLUP: Genomic- BLUP

PBLUP: Pedigree-BLUP

SE: Standard Error

MME: Mixed Model Equation

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Group records are the alternative solutions for traits that are affected by group interaction and are difficult or costly to measure on individuals. We aimed to investigate how group size, group composition, and validation sets affect the prediction accuracy using group records. Group records were the average of individual records in a tank. We used the genomic selection method and three validation sets; across-family, next-generation, and within-family. Group records (with tank effect) of a subfamily (size=20) gave prediction accuracy of 0.54, 0.62, and 0.74 for the above order validation sets. Instead, data without tank effect raised the above accuracies to 0.60, 0.67 and 0.78. However, group records of a family (size=40) had 0.01-0.03 lower accuracy than the subfamilies scenario. The group records based accuracy further declined sharply by 0.17-0.12 when a group comprised two unrelated families. Overall, we got 0.10-0.40 reduced accuracy using group records instead of individual ones. Hence, grouping by families, a higher number of group records, data without tank effect, smaller group sizes, and a close relationship between validation and reference populations improved genomic prediction accuracy using group records.

Keywords: Group records, Genomic selection, Accuracy, Across-family, Within-family.

1. INTRODUCTION

The world population is expected to grow to 9.15 billion by 2050 (Thornton 2010). It will increase the demand for animal protein and feed supplies. A more sustainable and suitable approach would be the selection and breeding of feed efficient animals since, feed shares up to 70% of total costs in livestock production (Alqaisi, Ndambi, and Williams 2017). Feed Conversion Ratio (FCR), the total amount of feed consumed per unit of production, measures the feed efficiency (De Verdal et al. 2018). In comparison to domestic animal species, aquaculture production has a high feed efficiency (de Verdal et al. 2018). But fish are reared in water and kept in large groups. The social interactions in large groups can affect the feed intake (Kooijman 2009). Thus, recording the feed intake of isolated fish might not represent the true feed intake in groups and is a very costly way of recording feed efficiency (every fish needs its own tank).

In animal and fish breeding, the major goal is to produce genetically improved animals in the future. Based on the estimates of breeding value (EBV) for a trait, parents are selected for breeding purposes. Researchers have developed different methods for estimating breeding values. The EBV estimation method is crucial for the accuracy of selection (Meuwissen, Hayes, and Goddard 2016). Meuwissen et al. (Meuwissen, Hayes, and Goddard 2001) introduced the "Genomic Selection" (GS) method which provides more accurate EBV of animals without records. In GS, a reference population is created, phenotyped and genotyped. Then, single nucleotide polymorphism (SNP) effects are estimated with the best linear, unbiased prediction (SNP-BLUP). The product of SNP effects with SNP genotypes will give genomic EBV for any newly genotyped individuals. Thus, phenotypic and genotypic records of the reference population are required to estimate the SNP effects. However, recording the feed efficiency at individual level is often difficult and expensive.

Group recording is one of the alternatives for the above problem. It is simple and reduces the cost of phenotyping every individual. For example, the total feed intake in a fish tank can be calculated by subtracting feed leftover from the feed administered to the tank. Then average feed intake per tank and average FCR per tank can be calculated. Similarly, the average genotype frequencies per tank can be computed by genotyping all the individuals in the tank or a pooled DNA sample from the tank. Then, we can use average records of the reference population to estimate the SNP effects. Now, individual

EBV can be predicted with SNP effects based on average records instead of individual records. This allows the selection of breeding individuals using group records.

2. OBJECTIVES

2.1. General Objective

We aim to predict the accuracy of genomic breeding value with SNP effects estimated from group records and compare them to estimates from individual records.

2.2. Specific Objectives

i. To develop genomic breeding value predictions based on group and individual records and apply them for within and across family predictions, and for the next generation.
ii. To assess the reduction in accuracy of breeding value estimates when group versus individual records are used.
iii. To assess the effects of group sizes and group compositions on the accuracy of breeding value estimates based on group records.

3. LITERATURE REVIEW

This literature review aims to discuss 1) Breeding Values and their estimates 2) Common methods for estimating EBV 3) EBV estimation with individual versus group records.

3.1. Breeding Value and their estimates (EBV)

In genetics, phenotypes are the observable characteristics of quantitative traits. The animal genotype and the type of environment in which animals grow affect their phenotype. Therefore, a part of observed phenotypic variation is because of the genotype variation. The genotypic variation is generally classified as additive genetic variation and non-additive genetic variation. The non-additive genetic variation is due to the interaction between alleles at the same locus (dominance effects) and alleles at different loci (epistatic effects). The non-additive genetic variation is often ignored because it does not add to the breeding value of an individual and it explains substantially less variance than the additive effects (Hill, Goddard and Visscher, 2008). The additive genetic variation is due to the additive effects of inherited alleles. In animal

breeding, this total additive genetic effect for a trait is known as the breeding value. Since the true breeding value of an animal is unknown, breeding values need to be estimated and are denoted as Estimated Breeding Values (EBV) (Su et al. 2012). The estimation is done by using own information and that of relatives. The degree of relationship among relatives determines the prediction accuracy of breeding value (Clark et al. 2012). Thus, relationships have been the fundamental unit in predicting breeding values.

3.2. Common methods for estimating EBV

Henderson (Henderson 1975) proposed a linear mixed model known as the Best Linear Unbiased Prediction (BLUP). BLUP has been used as the standard selection method in animal breeding. BLUP estimates the effect of a random variable or breeding value. BLUP includes the relationship information in estimating EBV which is important for traits that have low heritability (Sae-Lim et al. 2017). Based on the source of relatives' information, EBV estimation via BLUP can be broadly classified as follows:

3.2.1. Traditional BLUP:

In the traditional BLUP, the relationship among individuals is obtained by pedigree records (Mehrban et al. 2019). This method of EBV estimation using BLUP is known as Pedigree based BLUP (PBLUP).

PBLUP models the data as;

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Za} + \boldsymbol{e},$$

where $\boldsymbol{y}$ is the vector of phenotypic records, $\boldsymbol{b}$ is the vector of fixed effects, $\boldsymbol{a}$ is the vector of the additive genetic effects or breeding values, $\boldsymbol{e}$ is the vector of residual errors, and $\boldsymbol{X}$, $\boldsymbol{Z}$ are the incidence matrices linking $\boldsymbol{b}$ and $\boldsymbol{a}$ to $\boldsymbol{y}$, respectively. The random effects have the following distributions:

$$\boldsymbol{a} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{A}\sigma^2_{\boldsymbol{a}}),\ \boldsymbol{e} \sim \boldsymbol{N}(\boldsymbol{0}, \boldsymbol{I}\sigma^2_{\boldsymbol{e}})$$

where $\boldsymbol{A}$ is the additive genetic relationship matrix based on pedigree records, $\boldsymbol{I}$ is an identity matrix, and $\boldsymbol{\sigma^2_a}$, $\boldsymbol{\sigma^2_e}$ are additive genetic variance and residual error variances, respectively.

In PBLUP, we can select the animals only if phenotypic records are available on the animals themselves or their relatives. For example, in the dairy industry, the selection of

sires after the milk yields information from the daughter. The average age of the parents at the birth of their offspring is known as the generation interval. The annual genetic gain is low for longer generation interval species(De Roos et al. 2011). It further increases the costs of rearing progenies. Moreover, we cannot accurately select the animal until the progenies' performance is available in PBLUP.

3.2.2. Genomic BLUP

After 1990, DNA technology developed rapidly. Animal genetic value is better understood with DNA dissection rather than pedigree records. This grew the interest of scientists in selecting animals based on DNA information. For example: by Quantitative Trait Loci (QTL) mapping. QTL are mapped with the dense neutral markers exploiting linkage disequilibrium. It was assumed that the genes with significant effects affect the trait only. Consequently, many genes whose effect fails to reach the significance level were ignored. The significant genes were able to explain only ≤ 10% of genetic variation in animals. Thus, animal breeders lost interest in this approach (Meuwissen et al. 2016).

Meuwissen et al. (Meuwissen et al. 2001) introduced the "Genomic Selection (GS)" method. Unlike the previous methods, Genomic Selection assumes that every marker affects the trait. It estimates the effect of every SNP (single nucleotide polymorphisms) marker, known as the SNP effect. A reference population is created, phenotyped and genotyped to estimate SNP effects using the following BLUP model:

$$\boldsymbol{y} = \boldsymbol{Xb} + \boldsymbol{Zm} + \boldsymbol{e}\ ,$$

where $\boldsymbol{y}$, $\boldsymbol{X}$, $\boldsymbol{e}$ is as defined above, $\mathbf{m}$ is the SNP effects and $\boldsymbol{Z}$ is the matrix containing SNP genotypes of all individuals. The SNP effect has $\boldsymbol{m} \sim \boldsymbol{N}(\mathbf{0}, \boldsymbol{I\sigma^2_m})$ distribution, $\boldsymbol{e}$ has above defined distribution, and $\boldsymbol{\sigma^2_m}$ is SNP effect variance per marker. This method of estimating SNP effect is known as the SNP-BLUP method.

The SNP genotypes are coded as 0, 1, or 2 based on the allele counts. If a and A are two alleles, then 0=aa, 1=aA and 2=AA. The parameter estimates and the marker effects are independent of allele coding (Strandén and Christensen 2011).

In an alternative method, the SNP markers can be used to construct the genome-based relationship matrix ($\boldsymbol{G}$). If $\boldsymbol{Z}_{ij}$ is the original SNP genotype matrix coded as 0, 1 or 2 for

individual i (rows) at locus j (columns). Then $2p_j = \overline{\mathbf{Z}}_{j.}$ is the mean allele frequency or column average of $\boldsymbol{Z}_{ij}$. The ($\boldsymbol{G_{ij}}$) constructed from n SNP markers is calculated as:

$$\boldsymbol{G_{ij}} = \frac{(\boldsymbol{Z}_{ij} - 2p_j).(\boldsymbol{Z}_{ij} - 2p_j)'}{\sum_{j=1}^{n} 2p_j(1-p_j)}$$

where the denominator denotes the sum of heterozygosities.

The $(\boldsymbol{Z}_{ij} - 2p_j)$ is known as the centering of alleles. It is done to give more credit to rare alleles than to common alleles. Also, we get larger inbreeding coefficients if individuals are homozygous for rare allele (VanRaden 2008). Furthermore, the uncertainty of marker for estimating EBV is reduced by centering of alleles (Strandén and Christensen 2011). It is recommended to use base population allele frequencies. But since it is difficult to estimate base population allele-frequencies we often use $2p_j$ from the marker data. The scaling by $\sum_{j=1}^{n} 2p_j(1-p_j)$ is done to convert $\boldsymbol{G_{ij}}$ into a numerator relationship matrix (VanRaden 2008). The leading diagonal elements, $\boldsymbol{G_{ii}} = \mathbf{1} + \boldsymbol{Genomic\ Inbreeding\ coefficient}$ ($\boldsymbol{F_i}$). The $\boldsymbol{G}$ matrix then replaces the numerator relationship matrix ($\boldsymbol{A}$**)** of the PBLUP method. This method of calculating EBV is known as the GBLUP method.

The SNP effects can be back-solved from the EBV in GBLUP. It may be noted that the GBLUP and SNP-BLUP methods are equivalent, which implies that they are identical methods but parameterized differently and result in the same genomic EBV (VanRaden 2008).

After getting SNP effects from reference populations ($\hat{m}_j$), the EBV for any newly genotyped animal ($\hat{\boldsymbol{g}}_{\boldsymbol{i}}$) can be estimated as

$$\hat{\boldsymbol{g}}_{\boldsymbol{i}} = \sum_{j=1}^{n} \boldsymbol{Z}_{ij}.\hat{m}_j$$

where i, j, n and $\boldsymbol{Z}_{ij}$ is defined above. .

In the traditional methods, accurate EBV of an animal could only be calculated after the animal had its performance records or after its progenies had performance records. Thus, breeding programs required a long generation interval. Consequently, breeding progress was slow and annual genetic gains were small. But if an animal is genotyped

after birth or before birth, SNP- or G-BLUP may immediately yield accurate EBV. Hence, Genomic Selection (GS) often increases the annual genetic gain by 50-100% more than traditional breeding methods (Matthews et al. 2019). Furthermore, GS may give accurate EBV for traits that could not be measured directly on selected candidates like carcass traits. GS does not only increase the genetic gain, and provides better-estimated breeding value but, in dairy cattle, also saves millions of dollars associated with the progeny testing procedure. Large scale application of GS in the breeding industry was enabled by advanced, affordable, and high-throughput SNP genotyping technologies. All of these technologies revolutionized the breeding industries during the last decade (Meuwissen et al. 2016).

3.3. EBV estimation with individual versus group records

The SNP effect for a trait can be estimated from individual records as well as group records. Individual records are records on individual animals. Group records are pooled records or average records (Olson, Garrick and Enns, 2006). For example, a pond has two fish with individual records for the weight of 10 kg and 12 kg. The group record will be the sum of the individual records i.e., 22kg or the average of them i.e., 11kg. Now, EBV can be calculated with BLUP in two separate ways; either by using the group records or by using the individual records. The breeding value accuracy is defined as the correlation between the EBV and true breeding value (TBV). Furthermore, the breeding value accuracy depends on the number and information content of the records (Meuwissen, 2009). Individual records contain more information than group records as seen above. Thus, it is expected that the group records give less accuracy than individual records. But for some traits like feed intake group records are preferable to individual records, since they are more easily obtained and/or more abundant. Therefore, it is pivotal to know whether the use of group records can result in accurate EBV.

Olson et al. (Olson, Garrick and Enns, 2006) first proposed the use of group records or pooled observations with a PBLUP. The approach was applied by other researcher too (Cooper et al. 2010) (Peeters, Ellen, and Bijma 2013) (Biscarini, Bovenhuis, and Van Arendonk 2008). All of them included the additive genetic effect and a residual as a random effect only. Su et al. (Su *et al.*, 2018) used a multi-factored PBLUP model with the inclusion of the pen and litter effect as the random variables. Later, Ma et al. (Ma *et*

al., 2020) and Chu et al. (Chu *et al.*, 2019) used the GBLUP method for calculating EBV accuracy with group records and individual records. Most of these researchers used pooled/averaged observations as the group records.

The individual records-based Mixed Model Equations (MME) gives higher EBV accuracy than group records (Su *et al.*, 2018) (Chu *et al.*, 2019) (Ma *et al.*, 2020) (Biscarini et al. 2008). Similarly, when the pen effect was included in the MME, EBV accuracy with both individual and group records decreased (Ma *et al.*, 2020). Further, when the group size varies, EBV accuracy with group records also changes (Su *et al.*, 2018) (Olson, Garrick and Enns, 2006) . For example, a group size of 12 and 24 give prediction accuracy of 66.6% and 57.6% respectively (Su *et al.*, 2018). Thus, when group size increases and the number of groups decreases, the EBV accuracy from group records reduces. Similarly, the distribution of animals in pens also affects the EBV accuracy based on group records. When animals were allocated in a pen having a common sire, the EBV accuracy from group records was higher than when animals were allocated randomly (Olson, Garrick and Enns, 2006). Thus the selection of the appropriate group size and composition is essential for the breeding value accuracy with group records. Table 1 summarized the findings by different authors.

Table 1: Summary of literatures published on breeding value accuracy with individual and group records.

S.N.	Reference	Litter distribution	Group Size	GR	IR	GR/IR	Litter/Pen effect model
1.	Su et al., 2018	Common Sire/Parent	12	0.479	0.702	0.682	With
2.	Su et al., 2018	Common Sire/Parent	12	0.550	0.714	0.77	Without
3.	Ma et al., 2020	Common Sire/Parent	12	0.63	0.89	0.68	With
4.	Ma et al., 2020	Randomly distributed	12	0.55	0.89	0.62	With
5.	Ma et al., 2020	Common Sire/Parent	24	0.48	0.89	0.54	With
6.	Biscarini et al. ,2008	Common Sire/Parent	4	0.552	0.765	0.721	Without

7.	Biscarini et al., 200	Common Sire/Parent	4	0.461	0.759	0.607	Without
8.	Olson et al., 2006	Randomly distributed	2	0.50	0.63	0.793	Without
9.	Olson et al., 2006	Common sire	2	0.53	0.63	0.84	Without
10	Olsen et al (2006)	Randomly distributed	2	0.52	0.61	0.85	With
11	Olsen et al (2006)	Randomly distributed	12	0.21	0.63	0.33	Without
12	Olsen et al (2006)	Common sire	12	0.34	0.63	0.54	Without

GR: EBV accuracy with group records, IR= EBV accuracy with individual records

3.4. Conclusions from literature

Breeding animals are generally selected based on the EBV estimated from individual records. However, traits like feed efficiency are difficult and expensive to measure individually. But it can be easily measured at the group level reducing the phenotyping costs. Researchers have estimated the EBV and obtained a reasonable accuracy using group records. Most of them used the traditional BLUP for estimating EBV. Some researchers used the GBLUP model too. However, they focused on pooling the individuals and their phenotypic records. Genomic selection uses the SNP effects to estimate the EBV of an animal. The SNP effects for a trait are estimated from phenotypic and genotypic records of an individual. For a group recorded trait, pooled/averaged phenotypes and pooled/average genotypes can still be used to estimate the SNP effects. Therefore, we can apply the SNP-BLUP method to group records.

4. MATERIALS AND METHODS

This section is divided into the sections: i) statistical models for individual records, ii) statistical models for group records and iii) data simulation.

4.1. Statistical model for individual records

The linear mixed model equation for estimating SNP effects from individual records is given below.

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Zm} + \boldsymbol{Wt} + \boldsymbol{e}$$

Where $\boldsymbol{y}$ is a vector of phenotypic records; $\mathbf{1}$ is the vector of ones; μ is the unknown overall mean; $\boldsymbol{m}$ is a vector of SNP effects; $\boldsymbol{W}$ is a design matrix; $\boldsymbol{t}$ is a vector of tank effects and $\boldsymbol{e}$ is a vector of residual errors. If $\boldsymbol{X}_{ij}$ is the original SNP genotype matrix coded as 0, 1 or 2 for individual i at locus j and $2p_j$is the mean allele frequency from marker data, then a new SNP genotype matrix with centered alleles count ($\boldsymbol{Z}$) was calculated as (following VanRaden, 2008); $\boldsymbol{Z_{ij}} = \boldsymbol{X}_{ij} - 2p_j$

It was assumed that the random effects have distributions: $\boldsymbol{m \sim N(0, I.\sigma^2_m)}$, $\boldsymbol{t \sim N(0, I.\sigma^2_t)}$, $\boldsymbol{e \sim N(0, I.\sigma^2_e)}$,

where $\boldsymbol{\sigma^2_m}$, $\boldsymbol{\sigma^2_t}$, $\boldsymbol{\sigma^2_e}$ are variances of a marker effect, tank effect, and errors respectively. $\boldsymbol{I}$ is an identity matrix. The variance of the marker effects is:

$$\boldsymbol{\sigma^2_m} = \frac{Genotypic\ variance\ (G)}{sum\ of\ heterozygosities\ (H)} = \frac{Var(TBV)}{\Sigma 2p(1-p)}$$

4.2. Statistical Model for group records

The individual records based model can be transformed to a group records based model by:

$$\boldsymbol{Ty} = \boldsymbol{T}(\mathbf{1}\mu + \boldsymbol{Zm} + \boldsymbol{Wt} + \boldsymbol{e})$$

where $\boldsymbol{T}$ is a transformation matrix that links individual records to group records , and averages the records of a tank. If k be the number of individuals in a tank, $\boldsymbol{T}$ will transform k individual records to a single group record, i.e. the element $\boldsymbol{T_{ti}}$=1/k if animal i belongs to tank t and $\boldsymbol{T_{ti}}$=0 otherwise.

Then the model for group records is

$$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Z}^*\boldsymbol{m} + \boldsymbol{W}^*\boldsymbol{t} + \boldsymbol{e}^*$$

where $\boldsymbol{y}^* = \boldsymbol{T}\boldsymbol{y}$ is a vector of average records of the groups; $\boldsymbol{Z}^* = \boldsymbol{T}\boldsymbol{Z}$ is a SNP matrix of average allele (centered) counts of groups; $\boldsymbol{W}^* = \boldsymbol{T}\boldsymbol{W}$ is a matrix linking $\boldsymbol{t}$ to $\boldsymbol{y}^*$; $\boldsymbol{e}^* = \boldsymbol{T}\boldsymbol{e}$ is a vector of residual errors of group records.

The tank effect variance ($\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}}$) and SNP effect of a marker variance ($\boldsymbol{\sigma}_{\boldsymbol{m}}^{\mathbf{2}}$) are independent of the grouping of individuals. Thus $\boldsymbol{\sigma}_{\boldsymbol{t}}^{\mathbf{2}}$ and $\boldsymbol{\sigma}_{\boldsymbol{m}}^{\mathbf{2}}$ remain same for the group records. However, the group residual errors are the averages of residual errors of k individuals in a group. Since residual errors are independent and identically distributed with a constant variance $\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$;

$$Var(\boldsymbol{e}^*) = var(\boldsymbol{T}.\boldsymbol{e}) = var\left(\frac{1}{k}\sum_1^k e_i\right) = \frac{1}{\mathrm{k}}.\boldsymbol{\sigma}_{\boldsymbol{e}}^{\mathbf{2}}$$

Then, if $\hat{m}_j$ is the estimate of SNP effects from either individual or group records, genomic EBV for any newly genotyped individual ($\hat{\boldsymbol{g}}_{\boldsymbol{i}}$) can be obtained as

$$\hat{\boldsymbol{g}}_{\boldsymbol{i}} = \sum_{j=1}^{N_m} \boldsymbol{X}_{\boldsymbol{ij}}.\hat{m}_j$$

where N_m is the number of markers and i, j and $\boldsymbol{X}_{\boldsymbol{ij}}$ is defined as above.

The statistical models included in the study are given in Table 2.

Table 2: Statistical models analyzed in the study.

Records	Data with Tank effect	Data Without tank effect
Individual records	$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}\boldsymbol{m} + \boldsymbol{W}\boldsymbol{t} + \boldsymbol{e}$	$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Z}\boldsymbol{m} + \boldsymbol{e}$
Group records	$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Z}^*\boldsymbol{m} + \boldsymbol{W}^*\boldsymbol{t} + \boldsymbol{e}^*$	$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Z}^*\boldsymbol{m} + \boldsymbol{e}^*$

An example illustrating the individual records and group records model is presented in the Appendix. The data analysis was accomplished in Julia (Bezanson et al. 2012).

4.3. Data simulation

The data simulation represents the FCR values in a typical fish breeding scheme. QMSim generated the data (Sargolzaei and Schenkel 2009). The simulated heritability (h^2) estimate of 0.34 was assumed from FCR in Nile tilapia (De Verdal et al. 2018). The simulation of the historical population was done to create initial linkage disequilibrium (LD). A historical population size of unrelated 200 individuals was considered. No other evolutionary forces except genetic drift and mutation were allowed. The population undergoes random mating for 400 generations. Each generation consists of 200 individuals with equal numbers of males and females. Then 100 males and 100 females from this historical population were selected as the founder population. They were then randomly mated for 11 generations. Each pair of male and female would produce 50 offspring. Thus, each generation would have a total of 100 families and 5000 offspring.

For genomic information, there were five chromosomes each having a length of 100 cM. Each chromosome had 3000 markers and 100 QTLs (Quantitative Trait Loci), which were randomly positioned. Both QTL and markers were biallelic and have a recurrent mutation rate of $2.5*10^{-5}$. Their allele effects had a normal distribution. Thus, there were a total of 15000 genotype markers and 500 QTLs. The genotype file consists of animal identities and SNP genotype markers. Similarly, the animal data file has information on animal identity, phenotype, residual, and true breeding value (sum of QTL effects).

Each generation had data of 100 families, with 50 individuals per family. All data except generations 10 and 11 were excluded from the analysis. In generation 10, 90 out of 100 families were randomly chosen. The excluded 10 families (500 individuals) were across family validation sets. Then, 10 individuals per family were further excluded from each of the 90 chosen families. These 900 individuals were within family validation sets. The remaining 90 families (each having 40 individuals, i.e. a total of 3600 individuals) were the training data-sets. Similarly, from generation 11, all 5000 individuals were the next generation validation set.

The genotype file has the SNP genotypes coded as 0=a1a1, 2=a2a2, 3=a1a2, 4= a2a1. Here a1 is the paternal allele and a2 is the maternal allele. We transformed the above SNP code based on the allele frequency count as 0=a1a1, 1=a1a2=a2a1, 2=a2a2.

QMSim generated the normally distributed phenotypes. The phenotypic variance (σ^2_p) was 1. The genotype has a variance of 0.34 (σ^2_G) and the error variance (σ^2_e) was 0.66. Phenotypes here were taken as the without tank effect phenotypes. Next, we include the normally distributed tank effects with a variance of 0.3 (σ^2_t) to the original phenotypes. It increased the phenotypic variance to 1.3 (1+0.3). However, the genotypic variance and error variance remains the same as before. Both phenotypes with tank effect and without were analyzed in corresponding models (Table 2) accordingly.

In our study, group records are the average records of fish in a tank. The group sizes and fish composition in the tanks were varied to understand their effect on EBV estimation. For this, the scenarios in Table 3 were created.

Table 3: A description of the different scenarios. Each scenario was analysed using either individual or group records.

Scenarios	Group Composition	Group size (k)	Number of tanks (n)
I	Each individual fish in its own tank	1	3600
II	Each family was divided into two subfamilies and each subfamily was kept in a tank.	20	180
III	Each family in a tank	40	90
IV	Two unrelated families together in a tank	80	45
V	All unrelated families in a tank	3600	1

For each scenario, 20 replications were produced and analyzed. The genomic prediction accuracy was defined as the correlation between the genomic EBV and TBV. The prediction accuracy was calculated for within and across families and for the next generation validation set. The mean accuracies and standard error (SE) obtained from these 20 replicates are reported in the results.

5. RESULTS

Table 4, Table 5, and Table 6 showed the genomic prediction accuracy with standard errors for different validation sets. Each table presents results for data; with tank effect and without tank effect. The tables also contain the prediction accuracy of genomic EBV using group records and individual records. We sectioned the results as below.

Table 4: Next generation genomic prediction accuracy (mean ± SE of 20 replicates) based on group records and individual records in different scenario via data with the tank effect and without the tank effect.

Tank effect	Scenario	Group	Individual
Yes	I	0.838 ± 0.004	0.838 ± 0.004
	II	0.624 ± 0.009	0.811 ± 0.006
	III	0.597 ± 0.01	0.795 ± 0.007
	IV	0.455 ± 0.016	0.809 ± 0.006
	V	0	0.836 ± 0.005
No	I	0.863 ± 0.004	0.863 ± 0.004
	II	0.665 ± 0.009	0.863 ± 0.004
	III	0.636 ± 0.01	0.863 ± 0.004
	IV	0.512 ± 0.015	0.863 ± 0.004
	V	0	0.863 ± 0.004

Table 5: Within family genomic prediction accuracy (mean ± SE of 20 replicates) based on group records and individual records in different scenario via data with the tank effect and without the tank effect.

Tank effect	Scenario	Group	Individual
Yes	I	0.882 ± 0.003	0.882 ± 0.003
	II	0.737 ± 0.007	0.838 ± 0.006
	III	0.711 ± 0.008	0.820 ± 0.005
	IV	0.543 ± 0.013	0.840 ± 0.006
	V	0	0.880 ± 0.004
No	I	0.902 ± 0.003	0.902 ± 0.003
	II	0.776 ± 0.005	0.902 ± 0.003
	III	0.760 ± 0.006	0.902 ± 0.003
	IV	0.609 ± 0.01	0.902 ± 0.003
	V	0	0.902 ± 0.003

Table 6: Across family genomic prediction accuracy (mean ± SE of 20 replicates) based on group records and individual records in different scenario via data with the tank effect and without the tank effect.

Tank effect	Scenario	Group	Individual
Yes	I	0.819 ± 0.013	0.819 ± 0.013
	II	0.543 ± 0.031	0.757 ± 0.015
	III	0.532 ± 0.033	0.776 ± 0.015
	IV	0.383 ± 0.036	0.790 ± 0.013
	V	0	0.794 ± 0.016
No	I	0.833 ± 0.011	0.833 ± 0.011
	II	0.596 ± 0.024	0.833 ± 0.011
	III	0.583 ± 0.026	0.833 ± 0.011
	IV	0.461 ± 0.034	0.833 ± 0.011
	V	0	0.833 ± 0.011

5.1. Genomic prediction accuracy by using group records

Figure 1 shows the genomic prediction accuracy for different validation sets using group records with the tank effect. The prediction accuracy is highest in the scenario I, i.e. every fish in an individual tank. After that, the prediction accuracy from group records continuously drops with more individuals in scenarios II, III, IV and V, respectively. The reduction in prediction accuracy is substantial when different families are kept together in the tank (Scenario IV and Scenario V). Contrary to this, grouping individuals belonging to the same family like Scenario II and scenario III did not result in a pronounced decline in accuracy. For instance, the prediction accuracy declines by 0.01-0.03 from scenario II to scenario III and 0.14-0.17 from scenario III to scenario IV. We could not assess accuracy numerically in scenario V since there was only one record (all fish in one tank). In general, the prediction accuracy for group records reduces with increasing group size and decreasing relatedness.

Further, we noticed that the within-family selection gives most accuracy. The selection accuracy across the family is lower than within the family and next generation. The difference between within family and across family accuracy is widest in Scenario II, which is 0.20 but narrows down to 0.16 in scenario IV. Figure 2 shows similar results without tank effects and displays better accuracies than data with tank effects. For

example, the data without tank effects give 0.04-0.06 improved accuracy than data with tank effects in scenario II.

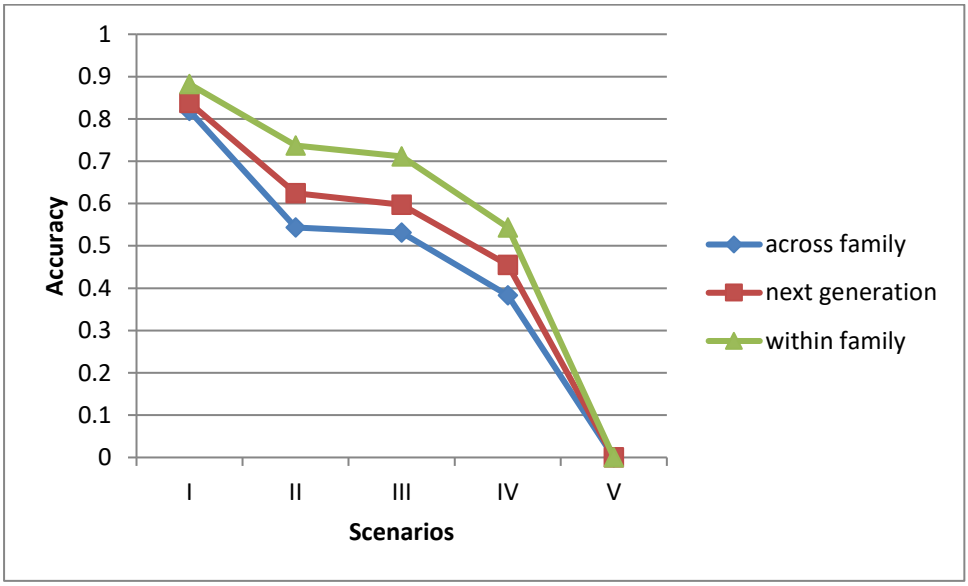

Figure 1: Genomic prediction accuracy (mean of 20 replicates) of EBV using group records for data with the tank effect across different scenarios and validation sets.

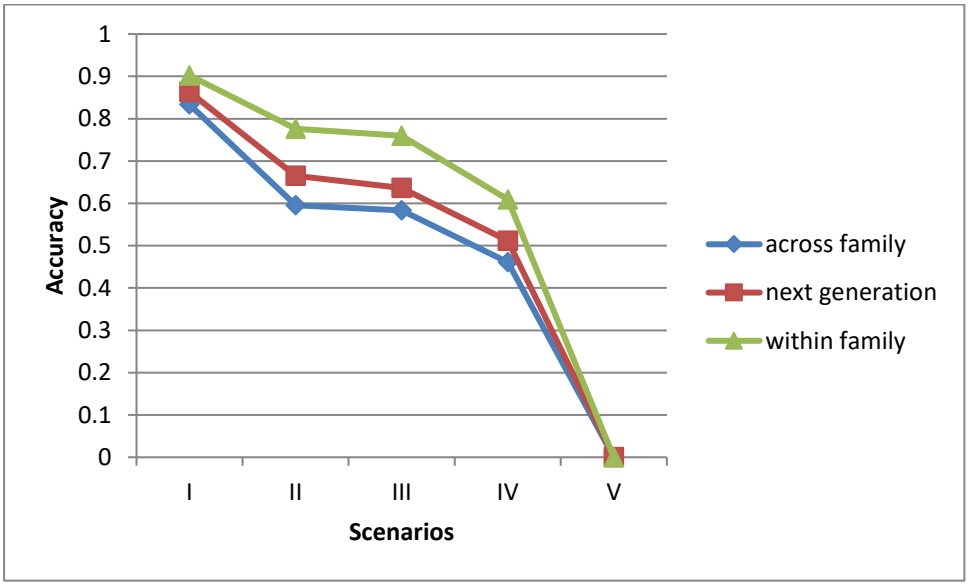

Figure 2: Genomic prediction accuracy (mean of 20 replicates) of EBV using group records for data without the tank effect across different scenarios and validation sets.

5.2. Genomic Prediction accuracy by using individual records

Figure 3 displays the genomic prediction accuracy for three validation sets using individual records with tank effect. As with group record based accuracy, the within-family validation sets give most accuracy, while the across family gives lowest accuracy. For within-family and next-generation validation sets, the genomic prediction accuracy gradually decreases from scenario II to scenario III by 0.016 and 0.018, respectively. Then, it rises from scenario III to scenario IV by 0.014-0.02. Scenarios I and V have similar accuracy for within-family and next-generation validation sets, which is also the highest accuracy using individual records. In the case of across families, the prediction accuracy increases from scenario II to scenario III by 0.014 and remains steady onwards.

Figure 4 illustrates the genomic prediction accuracy using individual records without the tank effect. It shows that within family accuracy is superior to next-generation/across family accuracy. However, the accuracies for each validation set remain constant throughout all scenarios, i.e. 0.90, 0.86, 0.83 for within family, next-generation and across family, respectively. Furthermore, the data without the tank effect gives better accuracy for individual records too.

Overall, the data with tank effect gives varying and lower accuracies, whereas the data without tank effect gives constant and higher accuracies while using individual records.

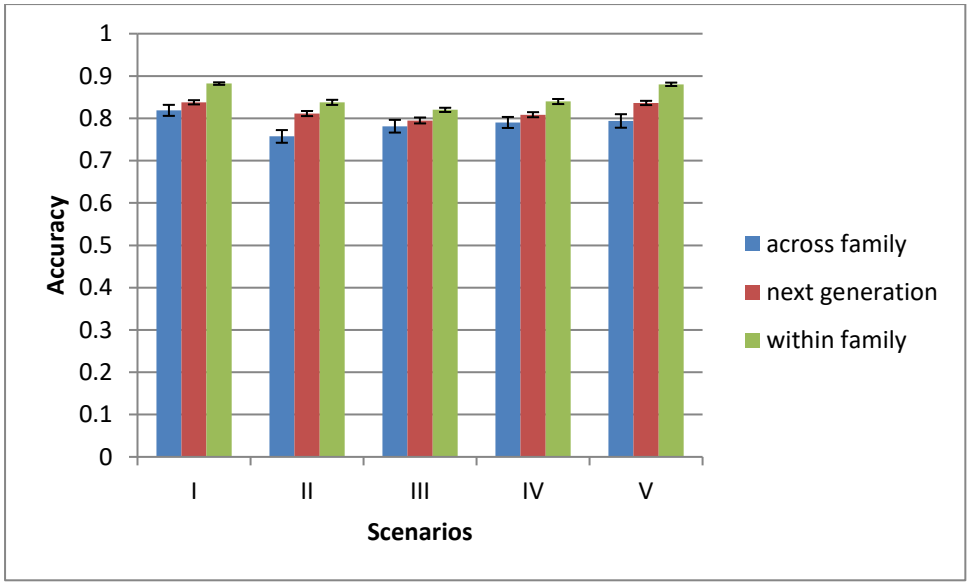

Figure 3: Genomic prediction accuracy (mean of 20 replicates) of EBV using individual records for data with the tank effect across different scenarios and validation sets.

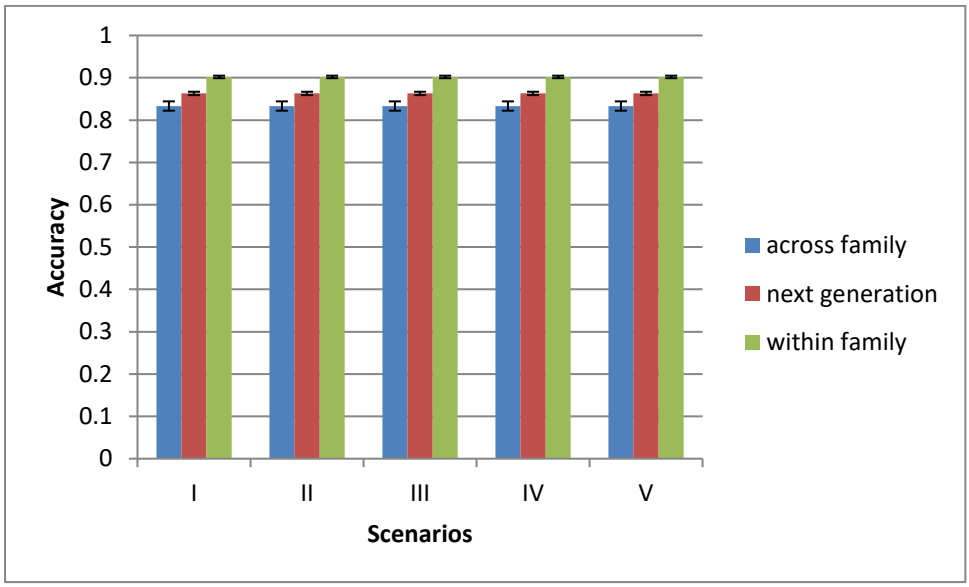

Figure 4: Genomic prediction accuracy (mean of 20 replicates) of EBV using individual records for data without the tank effect across different scenarios and validation sets.

5.3. Loss in genomic prediction accuracy using group records instead of individual records

Figure 5 illustrates the loss in genomic prediction accuracy for data with the tank effect, while figure 6 shows similar results without the tank effect. From both figures, we can conclude that across the family validation set has a considerable loss, whereas within the family has a low loss of accuracy. Among the three scenarios (II-IV), the loss is substantial in scenario IV, where two different families are together in a tank. For example, the accuracy loss in scenario IV for data with tank effect is about two times the accuracy loss in scenario II. The accuracy loss using group records is slightly higher in scenario III than in scenario II. The data without tank effect gives 0.02-0.04 more loss in accuracy in scenario II and scenario III than data with tank effect. However, in scenario IV, there is an almost similar loss in accuracy for data with tank effect and without tank effect, except across families. The accuracy loss is 0.04 more in data without tank effect than with tank effect in scenario IV for across family validation sets. Overall, the accuracy loss with group records instead of individual records ranges from 0.10 to 0.40.

We excluded scenario I and scenario V from figure 5 and figure 6 since they did not aid in comparing accuracy loss. Scenario I individuals records are the group records themselves, thus having no relative differences in their accuracies. Scenario V has only one group record, which gives a 100% accuracy loss when group records are used instead of individuals.

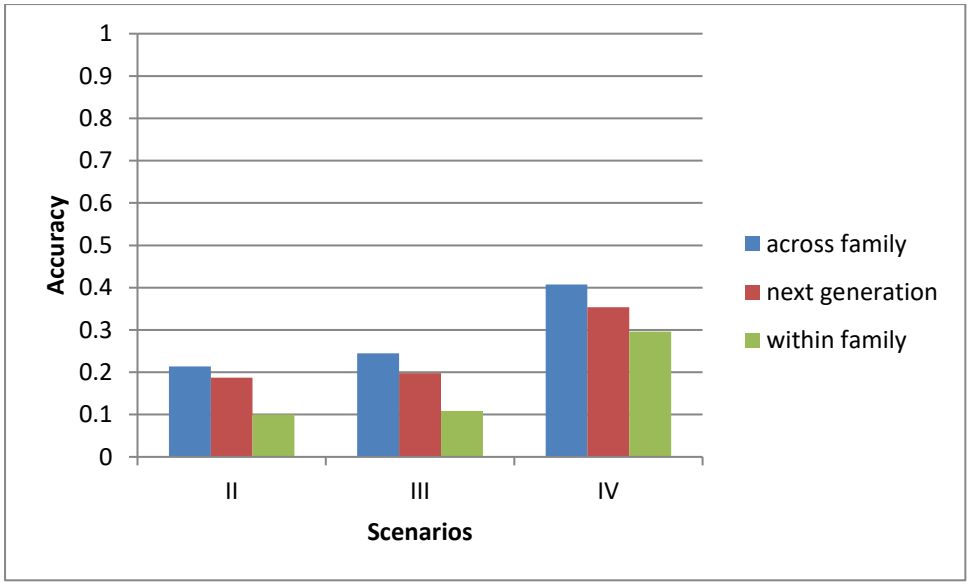

Figure 5: Loss in genomic prediction accuracy (mean of 20 replicates) in scenarios (II-IV) when group records were used instead of individual records for data with the tank effect across different validation sets.

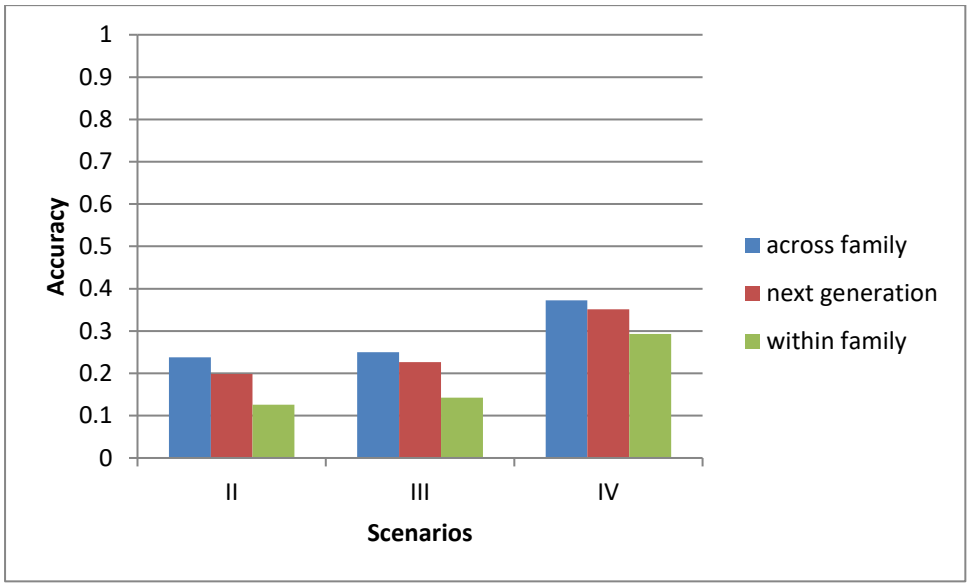

Figure 6: Loss in genomic prediction accuracy (mean of 20 replicates) in scenarios (II-IV) when group records were used instead of individual records for data without the tank effect across different validation sets.

5.4. Genomic prediction accuracy when individuals were randomly distributed in a tank.

Additionally, we randomly distributed the individuals in scenario III (ran_III) to understand the effect of animal distribution on prediction accuracies. We used data with tank effect only. Table 9 presents the animal distribution effect on genomic prediction accuracy. Group records based prediction accuracy declined almost to 0. The prediction accuracy based on individual records reduces across families by 0.031. But, accuracies from individual records increase slightly for within family and next-generation validation sets.

Table 7: Genomic prediction accuracy (mean ± SE of 20 replicates) when individuals in scenario III were randomly distributed (ran_III) using data with tank effect and different validation sets.

	Across Family		Next Generation		Within Family	
Scenarios	**Group**	**Individual**	**Group**	**Individual**	**Group**	**Individual**
ran_III	-0.073 ±0.039	0.745 ± 0.019	-0.018 ± 0.021	0.798 ± 0.008	-0.032 ±0.019	0.821 ±0.006
III	0.532 ± 0.033	0.776 ± 0.015	0.597 ± 0.01	0.795 ± 0.007	0.711 ± 0.008	0.820 ± 0.005

In general, our results showed that the group records, data with tank effect and across-family validation sets had higher standard errors than individual records, data without tank effect, and within-family/next generation validation sets. Similarly, a group with randomly distributed individuals or individuals from different families had a higher standard error than a group comprising family members.

6. DISCUSSION

6.1. With versus without tank effects:

The data without tank effects clearly gives higher genomic EBV accuracy than data with tank effect since the latter has higher phenotypic variation and lower heritability. Su et al. (Su et al., 2018) reported similar results for phenotypes with and without litter/pen effects.

6.2. Scenarios with group records:

Scenario I gives the highest accuracy. This is because there was no grouping of individuals, since every individual was in its own group. Thus, individual and group records-based accuracy remains the same in the scenario I. Olson et al.(Olson, Garrick and Enns, 2006) also penned everyone separately and did not categorize it as pooled or group records. In this sense, grouping individuals starts from scenario II. Where we took an average of each tank as our group records, the individual animal information is lost. The estimated SNP effect from group records was less accurate than the estimated SNP effect from individual records. Hence we found lower prediction accuracy by using group records instead of individual records. Ma et al. (Ma *et al.*, 2020), Su et al. (Su *et al.*, 2018), Chu et al. (Chu *et al.*, 2019), and Biscarini et al. (Biscarini et al. 2008) also compared the prediction accuracy based on individual records and group records. All of these studies reported reduced prediction accuracy from group records.

Our statistical model includes the tank effect (pen effect) but the group size is constant. Su et al. (Su *et al.*, 2018) suggested that the covariance matrix of the pen effect cannot be distinguished from the covariance matrix of the residual effect if the group size is constant. Thus, we cannot separate the tank effect and residual effect, which affects the reliability of prediction accuracy based on group records.

The genomic prediction accuracy increases when the number of training records (T) increases (Meuwissen, Hayes, and Goddard 2013) (Goddard 2009). Scenario I has more group-records than scenario II. The group composition remains constant in both scenarios. Hence, the fall in prediction accuracy was due to reduced number of records or increased group size. Our result matches with Su et al. (Su *et al.*, 2018), Ma et al. (Ma

et al., 2020) and Olson et al. (Olson, Garrick and Enns, 2006), who had shown that the prediction accuracy declines with increasing group size.

To understand the effect of group composition, we randomly distributed the individuals in scenario III. The group size remains constant. It gives a strong decline in prediction accuracy for group records. When a group contains related individuals, they have more similar genotypes than unrelated individuals. Then the average of each tank may more likely represent extreme genotypes, which increases the accuracy of estimation of SNP effects. Contrary to this, the estimated SNP effects from the group records of unrelated individuals (randomly distributed individuals) resemble each other more, i.e. each group is a random sample from the same population. If these random samples are sufficiently large (representative of the population), their mean performances and genotypes are expected to be very similar. Thus, when unrelated individuals are in a tank, we observe a decline in prediction accuracy. Olson et al. (Olson, Garrick and Enns, 2006) allocated the individuals in pens either based on common parent or randomly. The prediction accuracy declines in the latter case, which agrees with our results. Ma *et al.* (2020) also reported reduced genomic prediction accuracy for randomly distributed individuals in a tank.

For scenario IV, each tank has two unrelated families together. As explained above, the prediction accuracy from group records declines when unrelated individuals are present in the group. Also, there is only one single record in scenario V. Our group record is the average record itself. Then, the deviation of the group record from the average is 0. Hence all EBV are 0 in scenario V when using group records. Therefore, the prediction accuracy (numerically the correlation coefficient is not available) was 0.

6.3. **Individual records**:

We observed that the data without the tank effect gives constant accuracies across all scenarios, whereas the data with the tank effect provides varying accuracies. The reference population size and composition in a tank differ in the latter case. When we keep a family in a tank, we cannot separate the tank effect and family effect. This might be why we get reduced accuracy in scenario III. We can disentangle the family effect by including different families in a tank. For example, scenario IV and scenario V. Similarly, by splitting family members into separate tanks, helps estimating the tank effect. For

example, scenario I and scenario II. Likewise the random distribution of animals in scenario ran_III removes the spurious association between family effect and tank effect. Thus, we get improved accuracy for next generation and within family validation sets in scenario ran_III.

Contrary to this, across-family validation sets showed increased accuracy when the number of related individuals increased in a tank (from scenario II to scenario III). After scenario III, there is not any significant improvement in prediction accuracy as there is no addition of related individuals. Similarly, the random distribution of animals removes the relatedness of individuals in a tank. Thus, it might be the reason we obtained reduced accuracy when using across-family validation sets in scenario ran_III (Table 7).

6.4. **Within/across families and next generation validation**:

We found that genomic prediction accuracy within the families was higher than across families and next generation. Van den Berg et al. (van den Berg et al. 2019), Schopp et al. (Schopp et al. 2017), and Marjanovic et al. (Marjanovic and Calus 2021) found similar results for across and within family/breed analysis. The genomic prediction accuracy depends on the effective number of chromosome segments (M_e). The prediction accuracy increases if M_e decreases (Meuwissen et al. 2013) (Marjanovic and Calus 2021). There is a similar set of marker configurations between training and prediction sets for within family analysis. Hence, there is higher within family prediction accuracy.

M_e is greater across the families (van den Berg et al. 2019). Thus, the prediction accuracy declines across the families. Instead, we would require dense markers or whole genome sequences (WGS) to capture all associations between marker and QTL across the families (Marjanovic and Calus 2021). Meuwissen et al. (Meuwissen, van den Berg, and Goddard 2021) suggested Bayesian (Bayes) modelling for WGS. Van den Berg et al. (van den Berg et al. 2019) found that the Bayes R method yields higher accuracy than GBLUP, i.e. higher accuracy for across breed prediction (van den Berg et al. 2019). Therefore, high density markers or WGS and a non-linear variable selection models would be suitable for genomic prediction across families. This might be the reason why we did not find better genomic prediction accuracy across the families.

The next generation comprises offspring from the reference population and within/across families' validation sets. Thus, the next generation validation sets contain mixed individuals (related and unrelated to the reference population). Hence, the genomic prediction accuracy for the next generation is lesser than within family accuracy but higher than across family accuracy. Karaman et al. (Karaman et al. 2021) reported similar results using the mixed breed validation sets.

On the whole, it was expected that group records reduce accuracy relative to individual records. In this study, group records reduced accuracy by 0.10 -0.40. Whether this loss in accuracy is acceptable depends on the particular breeding programme: the costs of individual recordings relative to group recordings, the importance of the trait, the relationship between reference and validation sets, and whether grouping by families is possible. The results showed that grouping by families is very important to achieve accurate EBV and that having more/smaller groups improve the accuracy of genomic selection.

7. CONCLUSIONS

From estimated SNP effects based on group records we concluded

I. Genomic prediction accuracy of group records is highest when the groups contain related individuals.
II. Genomic prediction accuracy of group records is highest within the families, moderate in the next generation and lowest across the families.
III. Genomic prediction accuracy of group records reduces when the group size increases and the number of group records reduces.

In our study, scenario II, have higher genomic EBV accuracy than other scenarios across all validation sets. But this scenario requires more tanks and increased costs per family. It can be accommodated if we have a low number of families and our breeding goal is centered on group recorded traits. A scenario like IV provides better accuracy for individually recorded traits, reduced costs, and a reasonable accuracy for group recorded traits. Thus, it can be suited if we have larger families and are less interested in group recorded traits. Therefore, the choice of different scenarios largely depends on the significance of trait, group size, and relative cost of individual recording against group recording rather than on the type of validation sets.

REFERENCES

Alqaisi, Othman, Oghaiki Asaah Ndambi, and Ryan Blake Williams. 2017. "Time Series Livestock Diet Optimization: Cost-Effective Broiler Feed Substitution Using the Commodity Price Spread Approach." *Agricultural and Food Economics* 5(1):1–19.

van den Berg, I., T. H. E. Meuwissen, I. M. MacLeod, and M. E. Goddard. 2019. "Predicting the Effect of Reference Population on the Accuracy of within, across, and Multibreed Genomic Prediction." *Journal of Dairy Science* 102(4):3155–74.

Bezanson, Jeff, Stefan Karpinski, Viral B. Shah, and Alan Edelman. 2012. "Julia: A Fast Dynamic Language for Technical Computing."

Biscarini, F., H. Bovenhuis, and J. A. M. Van Arendonk. 2008. "Estimation of Variance Components and Prediction of Breeding Values Using Pooled Data." *Journal of Animal Science* 86(11):2845–52.

Chu, Thinh T., John W. M. Bastiaansen, Peer Berg, and Hans Komen. 2019. "Optimized Grouping to Increase Accuracy of Prediction of Breeding Values Based on Group Records in Genomic Selection Breeding Programs." *Genetics Selection Evolution* 51(1):1–12.

Clark, Samuel A., John M. Hickey, Hans D. Daetwyler, and Julius H. J. van der Werf. 2012. "The Importance of Information on Relatives for the Prediction of Genomic Breeding Values and the Implications for the Makeup of Reference Data Sets in Livestock Breeding Schemes." *Genetics, Selection, Evolution : GSE* 44(1):4.

Cooper, A. J., Calvin Ferrell, Larry V Cundiff, L. D. Van Vleck, A. J. ;. Cooper, Calvin ;. Ferrell, Larry V. ;. Cundiff, and Van Vleck. 2010. "Prediction of Genetic Values for Feed Intake from Individual Body Weight Gain and Total Feed Intake of the Pen." *Academic.Oup.Com* 88:1967–72.

Goddard, Mike. 2009. "Genomic Selection: Prediction of Accuracy and Maximisation of Long Term Response." *Genetica* 136(2):245–57.

Henderson, C. R. 1975. "Best Linear Unbiased Estimation and Prediction under a Selection Model." *Biometrics* 31(2):423.

Karaman, Emre, Guosheng Su, Iola Croue, and Mogens S. Lund. 2021. "Genomic

Prediction Using a Reference Population of Multiple Pure Breeds and Admixed Individuals." *Genetics Selection Evolution* 53(1):1–15.

Kooijman, S. A. L. M. 2009. "Social Interactions Can Affect Feeding Behaviour of Fish in Tanks." *Journal of Sea Research* 62:175–78.

Ma, Xiang, Ole F. Christensen, Hongding Gao, Ruihua Huang, Bjarne Nielsen, Per Madsen, Just Jensen, Tage Ostersen, Pinghua Li, Mahmoud Shirali, and Guosheng Su. 2020. "Prediction of Breeding Values for Group-Recorded Traits Including Genomic Information and an Individually Recorded Correlated Trait." *Heredity 2020 126:1* 126(1):206–17.

Marjanovic, Jovana and Mario P. L. Calus. 2021. "Factors Affecting Accuracy of Estimated Effective Number of Chromosome Segments for Numerically Small Breeds." *Journal of Animal Breeding and Genetics* 138(2):151.

Matthews, Daragh, John F. Kearney, Andrew R. Cromie, Fiona S. Hely, and Peter R. Amer. 2019. "Genetic Benefits of Genomic Selection Breeding Programmes Considering Foreign Sire Contributions." *Genetics Selection Evolution* 51(1):1–11.

Mehrban, Hossein, Deuk Hwan Lee, Masoumeh Naserkheil, Mohammad Hossein Moradi, and Noelia Ibáñez-Escriche. 2019. "Comparison of Conventional BLUP and Single-Step Genomic BLUP Evaluations for Yearling Weight and Carcass Traits in Hanwoo Beef Cattle Using Single Trait and Multi-Trait Models." *PLOS ONE* 14(10):e0223352.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. "Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps." *Genetics* 157(4):1819–29.

Meuwissen, Theo, Irene van den Berg, and Mike Goddard. 2021. "On the Use of Whole-Genome Sequence Data for across-Breed Genomic Prediction and Fine-Scale Mapping of QTL." *Genetics Selection Evolution* 53(1):1–15.

Meuwissen, Theo, Ben Hayes, and Mike Goddard. 2013. "Accelerating Improvement of Livestock with Genomic Selection." *Http://Dx.Doi.Org/10.1146/Annurev-Animal-031412-103705* 1:221–37.

Meuwissen, Theo, Ben Hayes, and Mike Goddard. 2016. "Genomic Selection: A Paradigm Shift in Animal Breeding." *Animal Frontiers* 6(1):6–14.

Meuwissen, Theo He. 2009. “Accuracy of Breeding Values of ‘unrelated’ Individuals Predicted by Dense SNP Genotyping.” *Genetics Selection Evolution* 41(1):1–9.

Olson, K. M., D. J. Garrick, and R. M. Enns. 2006. “Predicting Breeding Values and Accuracies from Group in Comparison to Individual Observations.” *Journal of Animal Science* 84(1):88–92.

Peeters, Katrijn, Esther Dorien Ellen, and Piter Bijma. 2013. “Using Pooled Data to Estimate Variance Components and Breeding Values for Traits Affected by Social Interactions.” *Genetics Selection Evolution* 45(1).

De Roos, A. P. W., C. Schrooten, R. F. Veerkamp, and J. A. M. van Arendonk. 2011. “Effects of Genomic Selection on Genetic Improvement, Inbreeding, and Merit of Young versus Proven Bulls.” *Journal of Dairy Science* 94(3):1559–67.

Sae-Lim, Panya, Antti Kause, Marie Lillehammer, and Han A. Mulder. 2017. “Estimation of Breeding Values for Uniformity of Growth in Atlantic Salmon (Salmo Salar) Using Pedigree Relationships or Single-Step Genomic Evaluation.” *Genetics Selection Evolution* 49(1):1–12.

Sargolzaei, Mehdi and Flavio S. Schenkel. 2009. “QMSim: A Large-Scale Genome Simulator for Livestock.” *Bioinformatics (Oxford, England)* 25(5):680–81.

Schopp, Pascal, Dominik Müller, Yvonne C. J. Wientjes, and Albrecht E. Melchinger. 2017. “Genomic Prediction within and across Biparental Families: Means and Variances of Prediction Accuracy and Usefulness of Deterministic Equations.” *G3: Genes, Genomes, Genetics* 7(11):3571–86.

Strandén, Ismo and Ole F. Christensen. 2011. “Allele Coding in Genomic Evaluation.” *Genetics Selection Evolution* 43(1):1–11.

Su, Guosheng, Ole F. Christensen, Tage Ostersen, Mark Henryon, and Mogens S. Lund. 2012. “Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers.” *PLoS ONE* 7(9).

Su, Guosheng, Per Madsen, Bjarne Nielsen, Tage Ostersen, Mahmoud Shirali, Just Jensen, and Ole F. Christensen. 2018. “Estimation of Variance Components and Prediction

of Breeding Values Based on Group Records from Varying Group Sizes." *Genetics Selection Evolution* 50(1):1–12.

Thornton, Philip K. 2010. "Livestock Production: Recent Trends, Future Prospects." *Philosophical Transactions of the Royal Society B: Biological Sciences* 365(1554):2853.

VanRaden, P. M. 2008. "Efficient Methods to Compute Genomic Predictions." *Journal of Dairy Science* 91(11):4414–23.

de Verdal, Hugues, Hans Komen, Edwige Quillet, Béatrice Chatain, François Allal, John A. H. Benzie, and Marc Vandeputte. 2018. "Improving Feed Efficiency in Fish Using Selective Breeding: A Review." *Reviews in Aquaculture* 10(4):833–51.

De Verdal, Hugues, Marc Vandeputte, Wagdy Mekkawy, Béatrice Chatain, and John A. H. Benzie. 2018. "Quantifying the Genetic Parameters of Feed Efficiency in Juvenile Nile Tilapia Oreochromis Niloticus." *BMC Genetics* 19(1):1–10.

APPENDIX

There are two tanks (n=2) each having three individuals (k=3) with phenotypic records of 10, 11, 12 (tank 1) and 7, 8, 9 (tank2) respectively. A 15000 SNP array (m=15000) with genotypes coded as 012 was used for genotyping.

Then, the model for individual records

$$\boldsymbol{y} = \mathbf{1}\mu + \boldsymbol{Zm} + \boldsymbol{Wt} + \boldsymbol{e}$$

$$\begin{bmatrix}\mathbf{10}\\ \mathbf{11}\\ \mathbf{12}\\ \mathbf{7}\\ \mathbf{8}\\ \mathbf{9}\end{bmatrix} = \begin{bmatrix}\mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\\ \mathbf{1}\end{bmatrix}\mu + \begin{bmatrix}(\mathbf{1}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{1}-\mathbf{1})\\ (\mathbf{0}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{2}-\mathbf{1})\\ (\mathbf{0}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{0}-\mathbf{1})\\ (\mathbf{1}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{0}-\mathbf{1})\\ (\mathbf{1}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{1}-\mathbf{1})\\ (\mathbf{0}-\mathbf{0.5}) \dots\dots\dots\dots (\mathbf{2}-\mathbf{1})\end{bmatrix}\begin{bmatrix}\boldsymbol{m_1}\\ .\\ .\\ .\\ .\\ \boldsymbol{m_{15000}}\end{bmatrix} + \begin{bmatrix}\mathbf{1}\ \mathbf{0}\\ \mathbf{1}\ \mathbf{0}\\ \mathbf{1}\ \mathbf{0}\\ \mathbf{0}\ \mathbf{1}\\ \mathbf{0}\ \mathbf{1}\\ \mathbf{0}\ \mathbf{1}\end{bmatrix}\begin{bmatrix}\boldsymbol{t_1}\\ \boldsymbol{t_2}\end{bmatrix} + \begin{bmatrix}\boldsymbol{e_1}\\ \boldsymbol{e_2}\\ \boldsymbol{e_3}\\ \boldsymbol{e_4}\\ \boldsymbol{e_5}\\ \boldsymbol{e}_{\mathbf{6}}\end{bmatrix}$$

The individual records transformed to group records as;

$$\boldsymbol{Ty} = \boldsymbol{T}(\mathbf{1}\mu + \boldsymbol{Zm} + \boldsymbol{Wt} + \boldsymbol{e})$$

$$\boldsymbol{T} = \begin{bmatrix}\frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\\ \mathbf{0}\ \mathbf{0}\ \mathbf{0}\ \frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}\frac{\mathbf{1}}{\mathbf{3}}\end{bmatrix}$$

Then the model for group records;

$$\boldsymbol{y}^* = \mathbf{1}\mu + \boldsymbol{Z}^*\boldsymbol{m} + \boldsymbol{W}^*\boldsymbol{t} + \boldsymbol{e}^*$$

$$\begin{bmatrix}\mathbf{11}\\ \mathbf{8}\end{bmatrix} = \begin{bmatrix}\mathbf{1}\\ \mathbf{1}\end{bmatrix}.\mu + \begin{bmatrix}-\mathbf{0.1666} \dots\dots\dots \mathbf{0}\\ \mathbf{0.1666} \dots\dots\dots \mathbf{0}\end{bmatrix}.\begin{bmatrix}\boldsymbol{m_1}\\ .\\ .\\ .\\ .\\ \boldsymbol{m_{15000}}\end{bmatrix} + \begin{bmatrix}\mathbf{1}\ \mathbf{0}\\ \mathbf{0}\ \mathbf{1}\end{bmatrix}.\begin{bmatrix}\boldsymbol{t_1}\\ \boldsymbol{t_2}\end{bmatrix} + \begin{bmatrix}\frac{\mathbf{1}}{\mathbf{3}}(\boldsymbol{e_1}+\boldsymbol{e_2}+\boldsymbol{e_3})\\ \frac{\mathbf{1}}{\mathbf{3}}(\boldsymbol{e_4}+\boldsymbol{e_5}+\boldsymbol{e_6})\end{bmatrix}$$

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

Postboks 5003
NO-1432 Ås
Norway