Accuracy of within-family multi-trait genomic selection models in a sib-based aquaculture breeding scheme

Binyam Dagnachew[a,b,*] and Theo Meuwissen[b]

[a] Nofima AS - Norwegian Institute of Food, Fisheries and Aquaculture Research, PB 210, N-1433 Ås, Norway

[b] Norwegian University of life Sciences, PB 5003, N-1433 Ås, Norway

* Corresponding author

Nofima AS - Norwegian Institute of Food, Fisheries and Aquaculture Research, N-1433 Ås, Norway

Email: binyam.dagnachew@nofima.no

Tel: +47 96729127

Abstract

Genomic selection can increase genetic gain in aquaculture breeding; however, its implementation is hindered by a high genotyping cost due to large individuals to genotype. Within-family genomic selection, which could utilize low-density markers and pedigree information, is suggested as a cost-effective way of implementing genomic selection in these species. In this study, a single trait genomic model (STGM) is compared with a multi-trait genomic model (MTGM) for prediction of within-family genomic breeding values in a simulated sib-evaluated aquaculture breeding scheme. Two traits, one with lower heritability (h_1^2=0.05) and another with higher heritability (h_2^2=0.5) were simulated. Three genetic correlations (r_g=0.2, r_g=0.5 and r_g=0.8) and zero residual correlation were assumed between these two traits. Given these assumptions, genomic and phenotypic data were simulated for 100 full-sib families of size 100. From each family, 10 individuals were randomly selected as selection candidates and the number of tested sibs varied from 10 to 90 per family. Two scenarios were investigated: in scenario I, all reference sibs were measured for both traits, whereas in scenario II half of the reference sibs measured for trait I and the remaining half were measured for trait II.

For both STGM and MTGM, prediction accuracies increased as the number of tested sibs per family increased from 10 to 90, however, the rate of increase was higher for STGM. Compared to STGM, use of MTGM increased the accuracy by up to 71% in scenario II and by up to 58% in scenario I for the low heritability trait when the genetic correlation between the traits was 0.8. The highest improvement in accuracy was observed in scenario II when only 10 sibs were genotyped per family with 10 SNP/Chr. As the magnitude of genetic correlation between the traits decreased, the relative gain in accuracy by implementing MTGM was reduced. The relative importance of MTGM also declined with the increase of number of tested sibs per family and a similar trend, but

with lesser magnitude, was observed with the increase of marker density. The results indicate that MTGM performs better than STGM for low heritability traits that are genetically correlated with high heritability traits. The advantage of multi-traits model was greater when both traits are not measured on the same group of individuals.

Keywords: aquaculture, genomic selection, within-family, single trait, multiple traits, sib-based

1. Introduction

Genomic selection (GS) is a novel method that uses genetic markers information for selecting parents of the future generation (Meuwissen et al., 2001). Currently, it is increasingly applied in livestock breeding programs particularly in dairy cattle (Goddard et al., 2007; Hayes et al., 2009; Goddard et al., 2010; Meuwissen et al., 2013). The benefits of GS are greater when traits of interest are not measured directly on selection candidates. In this regard, it is shown to have a very big potential in aquaculture breeding schemes because breeding goals in these species include many traits that are measured on sibs and not directly on the selection candidates (Sonesson, 2007; Sonesson et al., 2009; Sonesson et al., 2010; Odegard et al., 2014). A typical sib-based GS scheme in aquaculture involves estimation of marker effects in the sib of the candidates and the candidates are selected on breeding values estimated based on marker effects (Sonesson, 2007; Sonesson et al., 2009).

Application of conventional (full scale) GS in aquaculture species is very expensive due to the very large number of selection candidates and test-sibs to genotype. An alternative to overcome this particular challenge is to implement a combination of traditional BLUP for pre-selection of potential families and then estimate within-family genomic breeding values based on a low marker density (Lillehammer et al., 2013). This approach, referred as within-family (WF) genomic selection, can reduce genotyping cost without significant reducing selection accuracy because the low density markers can be used to trace inheritance within a family with a reasonable accuracy (Ødegård et al., 2014). Simulation studies confirmed that within-family genomic selection substantially improved prediction accuracies compared with conventional selection methods (Lillehammer et al., 2013; Ødegård et al., 2014).

71 Genetic correlations exist among traits included in many breeding goals and are indicators of
72 measurement from one trait carries information about other correlated traits. Prediction accuracies
73 could be improved by jointly evaluating these genetically correlated traits (Henderson et al., 1976;
74 Pollak et al., 1984; Schaeffer, 1984). The advantage of jointly modeling multiple traits compared
75 to analyzing each trait separately is that the inference process appropriately accounts for the
76 correlation among the traits, which helps to increase prediction accuracy and reduce trait selection
77 bias. In the context of genomic selection, studies reported that joint evaluation of multiple traits
78 benefits from genetic correlation between the traits and significantly improved prediction
79 accuracies (e.g. Calus and Veerkamp, 2011, Guo *et al*., 2014, Jiang *et al*., 2015). This is
80 particularly the case for lower heritability traits that are genetically correlated with higher
81 heritability trait. Jia and Jannink (2012) also reported that when phenotypes are not available for
82 all individuals and traits, better prediction accuracy is obtained for multiple traits genomic models
83 (MTGM) than for single trait genomic models (STGM).
84 Currently, within-family genomic selection models are tested only using single phenotype trait.
85 Therefore, the aim of this paper is to investigate the benefits of implementing multi-trait genomic
86 model in within-family genomic selection breeding schemes. Breeding schemes with different
87 number of tested sibs per family and different heritabilities of the traits under selection were
88 compared using computer simulation. In addition, different genetic correlation between traits was
89 investigated. Single and multi-trait genomic models were compared based on the accuracy of
90 selection.
91

92 ## 2. Methods

93 ### 2.1. Simulation of population and phenotypes

Datasets were simulated to compare single and multi-trait within-family genomic selection methods. A genomic data was simulated assuming an effective population size of 1000 (N_e), 500 males and 500 females, and was kept for 4000 generations to achieve mutation-drift-LD balance. The simulated genome consisted of 10 chromosomes each 1M (Morgan) length and 11,000 bi-allelic loci across the genome. At generation 4001, ~ 1,100 marker loci and ~ 30 QTL segregated with minor allele frequency (MAF) of above 5% at each of 10 chromosomes (i.e. a total of 1,100 per chromosome). In 4001 generation, a pedigree structure of 50 males and 100 females was used to create 100 full-sib families of family size 100 giving 10150 individuals including the base generation. A schematic description of the simulation is presented in Figure 1.

2.2. Data structure

True breeding values were obtained as the sum of all QTL additive effects for each individual. Phenotypes were defined as the sum of true breeding values and random residuals sampled form standard normal distribution. Two traits were simulated: Trait I with heritability $h_1^2 = 0.05$ and Trait II with heritability $h_2^2 = 0.5$. Three datasets were generated assuming different genetic correlation between Trait I and Trait II. The simulated genetic correlations were $r_g = 0.2$, $r_g = 0.5$ and $r_g = 0.8$ and the residual correlation between the two traits was assumed zero.

2.3. Marker density

Four different marker densities containing 10, 20, 50 and 100 marker per chromosome were generated by uniformly sampling markers from the complete dataset. The within-family genomic relationship matrix (G) was constructed based on linkage analysis (Luan et al., 2012). Genotype inheritance probabilities were estimated using Linkage Disequilibrium Multi-locus Iterative Peeling (LDMIP) program (Meuwissen et al., 2010) based on information from markers and the pedigree. The output was then used to calculate the genome-wide relationship matrix.

2.4. Breeding value estimation

Breeding values were estimated as a combination of family breeding values and genomic within-family breeding values. $EBV = 1/2\, a_s + 1/2\, a_d + u$, where EBV was a vector of combined breeding values, a_s and a_d were vectors of conventional BLUP estimated breeding values of the sire and dam of an individual respectively, and u was the within-family genomic breeding values of individuals.

Family breeding values were estimated using conventional BLUP methodology. Two different models were used to estimate within-family genomic breeding values. The single trait genomic model (STGM) was:

$$y = \mu + Zu + e,$$

Where y is a vector of phenotypes, μ is the overall mean, Z is a design matrix linking animals to the observation, u is a vector of estimated within-family genomic breeding values and e is a vector of random residuals. It is assumed that $u \sim N(0, \frac{1}{2} G\sigma_u^2)$, where G is a genomic relationship matrix for the animals in a full-sib family and σ_u^2 is the additive genetic variance; and $e \sim N(0, I\sigma_e^2)$, where **I** is an identity matrix and σ_e^2 is residual variance. The G was calculated based on linkage analysis performed using the LDMIP program (Meuwissen et al., 2010). This method uses an iterative peeling step for each genotype locus to account for family information. The G matrix was calculated for each full-sib family.

The general form of the multi-trait genomic models (MTGM) was:

$$\begin{bmatrix} y_1 \\ y_2 \end{bmatrix} = \begin{bmatrix} 1 & 0 \\ 0 & 1 \end{bmatrix} \begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix} + \begin{bmatrix} Z_1 & 0 \\ 0 & Z_2 \end{bmatrix} \begin{bmatrix} u_1 \\ u_2 \end{bmatrix} + \begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$$

Where $\begin{bmatrix} y_1 \\ y_2 \end{bmatrix}$ is a vector of phenotypes for traits I and II, 1 is a vector of ones, $\begin{bmatrix} \mu_1 \\ \mu_2 \end{bmatrix}$ is a vector of overall means for traits I and II, Z_1 and Z_2 are design matrices linking animals to the observation,

139 $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix}$ is a vector of estimated within-family genomic breeding values for the two traits, and $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix}$ is

140 a vector of random residuals for the two traits. It is assumed that $\begin{bmatrix} u_1 \\ u_2 \end{bmatrix} \sim$

141 $N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \frac{1}{2}\begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_{12}} \\ \sigma_{u_{21}} & \sigma^2_{u_2} \end{bmatrix} \otimes G)$, where $\begin{bmatrix} \sigma^2_{u_1} & \sigma_{u_{12}} \\ \sigma_{u_{21}} & \sigma^2_{u_2} \end{bmatrix}$ is the additive genetic variance and covariance

142 structure; and $\begin{bmatrix} e_1 \\ e_2 \end{bmatrix} \sim N(\begin{bmatrix} 0 \\ 0 \end{bmatrix}, \begin{bmatrix} \sigma^2_{e_1} & 0 \\ 0 & \sigma^2_{e_2} \end{bmatrix} \otimes I)$, where $\begin{bmatrix} \sigma^2_{e_1} & 0 \\ 0 & \sigma^2_{e_2} \end{bmatrix}$ is the residual variance and

143 covariance structure. Other model components are as defined previously. For both STGM and

144 MTGM, the within-family genomic breeding values were predicted from a single trait and multi-

145 trait models in WOMBAT (Meyer, 2007), respectively. The genomic relationship matrix, G,

146 calculated based on the marker and pedigree information using LDMIP was fit in WOMBAT.

147 **2.5. Scenarios compared**

148 Two different scenarios were compared to investigate the performance of the models to predict

149 within-family genomic breeding values. In scenario I, all tested sibs had phenotypes for both traits

150 and in scenario II, half of the tested-sibs had phenotype for trait I and the other half had phenotype

151 for the second traits. Scenario II emulates a practical situation where some group of sibs of

152 selection candidates are challenged for a certain disease and the remaining sibs are used to obtain

153 measurements for other traits. Furthermore, the effect of the number of genotyped and phenotyped

154 sibs was investigated by varying the number of tested sibs per family (animals with both genotype

155 and phenotype). Of the 100 sibs in each family, 10 sibs were chosen randomly as selection

156 candidates (non-phenotype validation animals) and the number of tested sibs varied from 10 to 90.

157 Each scenario was also tested under different marker densities.

158 **2.6. Criteria of comparison**

159 The predictive abilities of STGM and MTGM was investigated by masking the phenotypes of 10
160 randomly selected candidates from each family (validation sibs) and predict their breeding values.
161 The evaluation was based on 30 replicates for each tested scenario and the average of the replicates
162 was reported. The performance of STGM and MTGM were evaluated using the accuracy of
163 prediction and the bias of the estimates. Accuracy of prediction was calculated as the Pearson's
164 correlation between true (i.e. simulated) and estimated breeding values.
165

166 ## 3. Results

167 ### 3.1. Effect of family size and heritability

168 The effect of family size on prediction accuracy was tested by varying the number of tested sibs
169 from 10 to 90 per family. Table 1 presents prediction accuracies of STGM and MTGM under
170 scenarios I and II for marker density of 100 SNP/Chr (the results of the other marker densities are
171 not presented here, but similar trends are observed). For the STGM, prediction accuracies for both
172 traits are reported, however, for the MTGM, prediction accuracies for only trait I is reported for
173 the three genetic correlations between the two traits. Across both scenarios and heritabilities, the
174 prediction accuracy increased as the number of tested sibs per family increased from 10 to 90
175 (Table 1). The table also shows the effect of family size on prediction accuracy is less pronounced
176 on the higher heritable trait compared to the lower heritable trait. The effect of number of tested
177 sibs on accuracy was less on MTGM compared to STGM. For instance, in scenario I, for trait I,
178 the increase in accuracy was 69% under STGM as the size of tested sibs increased from 10 to 90,
179 however, for the MTGM it was 63%, 42% and 28.5% when the genetic correlation between the
180 traits was 0.2, 0.5 and 0.8 respectively. A similar trend but more pronounced effect of family size
181 was observed for trait I in scenario II (Table 1). The highest increase in accuracy was observed (up

to 78%) for the lower heritability trait (i.e. trait I) in scenario II for STGM when the number of tested sibs increased from 10 to 90 (Table 1).

3.2. Effect of marker density

Four marker densities, 100, 50, 20 and 10 SNP/Chr were studied to test the effect of marker density on prediction accuracies of WF breeding values. Figure 2 plots the prediction accuracies for trait I (h^2=0.05) and trait II (h^2=0.5) based on STGM and MTGM for scenario I. The figure shows that the lower heritability trait under STGM is more sensitive to the marker density and to family size than the higher heritability trait (Figure 2 top right). It is also showed that compared to STGM (Figure 2 top right), the use of multi-trait model has reduced the sensitivity to marker density for trait I (Figure 2 bottom right). However, less marker density sensitivity was observed for trait II for both STGM and MTGM (Figure 2 left top and bottom).

3.3. Effect of genetic correlation

Three genetic correlations between traits (i.e. 0.2, 0.5 and 0.8) were tested to investigate the effect of it on the prediction ability of MTGM. Figures 3 and 4 present relative gain in accuracy for using MTGM instead of STGM for trait I under scenario I and II, respectively. In general, as the genetic correlation between the two traits increased, the relative gain in accuracies also improved for the lower heritability trait (Figures 3 and 4). When the genetic correlation between trait I and II was 0.2, the gain in accuracy for trait I was under 6% in scenario I (Figure 3) and under 13% in scenario II (Figure 4). However, when the genetic correlation increased to 0.5, the relative gain in accuracy increased up to 29% in scenario I (Figure 3) and up to 37% in scenario II (Figure 4). The accuracy of prediction for trait I was improved by 58% in scenario I (Figure 3) and by 71.2% in scenario II (Figure 4) by using MTGM when the genetic correlation between trait I and II was increased to 0.8. There is little or no relative benefit in accuracy was observed by using MTGM in place of

205 STGM for trait II (results not presented). Regardless of the marker density, the relative gain in
206 accuracy for using MTGM decreased as the number of tested sibs increased from 10 to 90 per
207 family and the extra gain beyond family size of 60 was minimal in both scenarios.

208 **3.4. Effect of missing phenotype**

209 In scenario I all the test sibs were measured for the two traits (i.e. trait I and II), whereas in scenario
210 II one half of the test sibs were measured for trait I and the remaining half of sibs were measured
211 for trait II. The relative importance of MTGM was greater in scenario II, where not all tested sibs
212 are phenotyped for both traits. Up to 71.2% relative gain in accuracy was observed in scenario II
213 (Figure 4) compared to up to 58% in scenario I (Figure 3) for trait I when the genetic correlation
214 between the traits was 0.8. As the genetic correlation between traits decreased, the relative
215 importance of MTGM also reduced. The highest relative gain in accuracy by using MTGM was
216 observed when marker density was 10 SNP/Chr and only 10 sibs per family were tested in scenario
217 II (Figure 4). When marker density increased from 10 SNP/Chr to 100 SNP/Chr, the relative gain
218 in accuracy for using MTGM decreased gradually in both scenarios. For instance, in scenario I
219 when only 10 sibs are tested, the % gain in accuracy was 58 for 10 SNP/Chr and it reduced to 48.4
220 when the marker density increased to 100 SNP/Chr (Figure 3). The reduction in % of gain in
221 accuracy was from 71.2 to 55.7 in scenario II when the marker density increased to 100 SNP/Chr
222 (Figure 4). Furthermore, it is also observed that as the number of tested sibs per family increased
223 from 10 to 90, the relative importance of MTGM over STGM decreased in both scenarios (Figures
224 3 and 4).

225

226 **4. Discussion**

Most of existing aquaculture breeding schemes are family-based where within-family additive genetic variation is hardly exploited for selection. Thus, selection methods allowing within-family selection are of importance. Within-family genomic selection is a strategy where genomic information is used to account for within-family variation component (Lillehammer et al., 2013; Ødegård et al., 2014). It is complementary with the existing family selection breeding schemes where the between family component of the breeding values are estimated by pedigree based method (Lillehammer et al., 2013). Studies reported that a substantial increase in accuracy obtained through implementation of WF genomic selection based on sparse marker density (Lillehammer et al., 2013; Ødegård et al., 2014). Sparse marker densities are sufficient in WF genomic selection because of the large family sizes in fish breeding. Previous studies have compared the prediction accuracies of within-family genomic breeding values using single phenotypes. In the current study, we have compared the performance of STGM and MTGM to predict the WF genomic breeding values.

The benefit of multi-trait models over single trait models comes from the fact that it uses an extra information from genetically correlated traits. Genetic correlation is a key factor determining the advantage of multi-trait models over single trait models. The current study showed that MTGM improved prediction accuracies for the lower heritability trait, which was genetically correlated with higher heritable trait (Table 1, Figures 3 and 4). The importance of MTGM was higher when not all test individuals measured for the two traits studied (Figure 4). Guo et al., (2014), reported that more accurate breeding values were obtained with MTGM than STGM for traits that had missing data and are genetically correlated with higher heritability trait. The level of genetic correlation determined the degree of improvement obtained from MTGM. Figures 3 and 4 showed that with lower genetic correlation (i.e. r_g= 0.2) the gain from MTGM was minimal (under 6% and

13% for scenario I and II respectively). This is in agreement with results reported in other studies (Calus et al., 2011; Jia et al., 2012; Guo et al., 2014; Jiang et al., 2015).

The difference between scenario I and II is that in scenario I all test sibs are phenotyped for both traits, whereas in scenario II only half of the test sibs were phenotyped for trait I and the other half were phenotyped for trait II. The current study showed that the benefit of MTGS was higher in scenario II (Table 1, Figures 3 and 4). Previous studies have also reported a greater advantage of multi-trait models for lower heritability traits that had missing data and are genetically correlated with higher heritability traits (Hayashi et al., 2013; Guo et al., 2014). In practical aquaculture breeding programs, phenotype measurements for all traits of interest are not often available for all test sibs. For example, in a typical sib-based breeding program, only subset of test sibs are phenotyped for traits that are difficult or expensive to measure such as filet quality and disease resistance traits. Accuracies obtained using STGM were lower for trait I especially when the number of tested sibs per family was under 50 (Table 1). These accuracies have improved greatly under MTGS by using information from the genetically correlated higher heritability trait (i.e. trait II with rg = 0.5 and rg = 0.8, Table 1). However, for the higher heritability trait, MTGS made no substantial difference in prediction accuracy as it is also reported in (Hayashi et al., 2013; Guo et al., 2014).

In aquaculture breeding, considerable weight in the breeding goal is put on disease resistance traits. These traits are, however, expensive to measure and some of them have low to medium heritability (Guy et al., 2009; Drangsholt et al., 2011; Lhorente JP, 2014) hence are challenging to improve through traditional selective breeding. On the other hand, production traits such as growth rate, carcass yield and fillet yield have higher heritability (e.g. Rye and Refstie, 1995, Powell et al., 2008) and selection is relatively more effective for these traits. Sonesson and Meuwissen (2009)

and Nielsen et al., (2009) observed that use of genomic selection could considerably increase accuracy of selection in aquaculture species, particularly for traits that are difficult to measure on selection candidates themselves, for instance, disease resistance traits. Lillehammer et al. (2013) presented WF genomic selection, a cost-effective implementation of GS in aquaculture breeding using low-marker density. The current study showed that the accuracy of WF genomic selection from single phenotype could be substantially improved by including multiple phenotypes in the genetic evaluation, particularly for lowly heritable traits (Figures 3 and 4). Genetic correlations exist between resistance against different diseases (e.g. in Atlantic salmon Gjøen et al., 1997) and between some disease resistance traits and harvest body weight traits (e.g. in Atlantic salmon Drangsholt et al., 2012). These correlations could be exploited in multi-trait genomic models to improve prediction accuracies in aquaculture breeding programs.

Increasing marker density is expected to increase prediction accuracy of genomic breeding values (e.g., Solberg et al., 2008; Nielsen et al., 2009). The current study showed that marker density has a small effect on WF genomic breeding values prediction accuracies (Figure 2). This is in agreement with previous reports (Lillehammer et al., 2013; Ødegård et al., 2014; Ødegård et al., 2015). The impact of marker density is small because genomic relationship matrices are constructed within full sib families (i.e. equivalent to an effective population size of 2) and few markers are adequate to trace inheritance. However, the effect of marker density was stronger on the lower heritability trait than the higher heritability trait (Figure 2). The current study also found that the accuracy of selection increased as the number of genotyped sibs per family increased (Table 1). The relative gain in accuracy for using MTGM, however, decreased as the number of tested sibs increased from 10 to 90 per family and the extra gain beyond family sizes of 60 was

minimal (Figures 3 and 4). Nirea et al. (2014), also reported that genotyping more than 60 per family yields relatively little added value.

The economic efficiency of WF genomic selection relies on two aspects of the design; pre-selection of families to reduce genotyping cost and use of sparse dense marker to construct the within-family genomic relationship matrix (Lillehammer et al., 2013; Ødegård et al., 2014; Ødegård et al., 2015). However, unlike conventional genomic selection programs in other species, where reference population can be re-used, it requires re-building of reference population every generation. Consequently, if obtaining phenotype for a trait is expensive or difficult, WF genomic selection will have more challenges compared to conventional genomic selection. Hence, if phenotyping is limiting, as in the case of scenario II, analyzing genetically correlated traits together is more beneficiary WF genomic selection (Figure 3).

The current study compared MTGS and STGS models for prediction of WF breeding values using a linear model (GBLUP) under a single genetic architecture. In previous studies, linear and non-linear (Bayesian) multi-trait models were compared under a single genetic architecture (Calus et al., 2011; Jiang et al., 2015) and multiple genetic architectures (Jia et al., 2012; Montesinos-López et al., 2016). They reported that GBLUP gave relatively consistent performance across different genetic architecture and under a major QTL genetic architecture, the Bayesian models performed better than GBLUP in both single and multi-trait models. It is also reported that MTGS was strongly beneficial under a major QTL genetic architecture than under a polygenic genetic architecture (Jia et al., 2012). Hence, if a trait of interest is known to be affected by major genes, for instance, resistance to infectious pancreatic necrosis (IPN) in Atlantic salmon (Houston et al., 2008; Moen et al., 2009), implementation of multi-trait non-linear models could be considered.

5. Conclusion

Results show that a MTGM can improve accuracy of within-family genomic prediction for lower heritability traits that are genetically correlated with higher heritability traits. The importance of multi-traits models were greater when both traits are not measured on the same group of individuals. On the other hand, there is little or no improvement in accuracy by choosing MTGS over STGM when the genetic correlation between traits is low. The prediction accuracy of within-family breeding values increased as the number of tested sibs per family increased. The relative importance of MTGS over STGM, however, decreased as family size increased and was minimal beyond test sib size 60. Thus, when resources are limiting, genotyping 60 individuals per family would obtain a substantial benefit.

Figures

Figure 1: A diagrammatic illustration of the simulated structure of the population

Figure 2: Prediction accuracies of the two traits with heritabilites of 0.5 and 0.05 and different marker densities in Scenario I. For the MTGM, the genetic correlation between Trait I and II was 0.8.

Figure 3: Relative gain in accuracy in percentage by using MTGM instead of STGM for prediction of within-family genomic breeding values with different marker density under Scenario I

Figure 4: Gain in prediction accuracy in percentage by using MTGM instead of STGM for prediction of within-family genomic breeding values with different marker density under Scenario II

Tables

Table 1: Prediction accuracies and standard error of STGM and MTGM under scenarios I and II. For the MTGM, prediction accuracies for only trait I is reported for the three genetic correlations between the two traits. The marker density is 100 SNPs/Chr.

CRediT author statement

Binyam Dagnachew: Conceptualization, Methodology, Writing- Original draft preparation, Writing- Reviewing and Editing. **Theo Meuwissen:** Conceptualization, Supervision, Writing- Reviewing and Editing

6. References

Calus M.P., Veerkamp R.F., 2011. Accuracy of multi-trait genomic selection using different methods. Genetics Selection Evolution. 43, 26.

Dominik S., Henshall J.M., Kube P.D., King H., Lien S., Kent M.P., et al., 2010. Evaluation of an Atlantic salmon SNP chip as a genomic tool for the application in a Tasmanian Atlantic salmon (Salmo salar) breeding population. Aquaculture. 308, Supplement 1, S56-S61.

Drangsholt T.M.K., Gjerde B., Odegard J., Finne-Fridell F., Evensen O., Bentsen H.B., 2011. Quantitative genetics of disease resistance in vaccinated and unvaccinated Atlantic salmon (Salmo salar L.). Heredity. 107, 471-7.

Drangsholt T.M.K., Gjerde B., Ødegård J., Finne-Fridell F., Evensen Ø., Bentsen H.B., 2012. Genetic correlations between disease resistance, vaccine-induced side effects and harvest body weight in Atlantic salmon (Salmo salar). Aquaculture. 324–325, 312-4.

Gjøen H.M., Refstie T., Ulla O., Gjerde B., 1997. Genetic correlations between survival of Atlantic salmon in challenge and field tests. Aquaculture. 158, 277-88.

Goddard M.E., Hayes B.J., 2007. Genomic selection. Journal of animal breeding and genetics = Zeitschrift fur Tierzuchtung und Zuchtungsbiologie. 124, 323-30.

Goddard M.E., Hayes B.J., Meuwissen T.H., 2010. Genomic selection in livestock populations. Genetics research. 92, 413-21.

Guo G., Zhao F., Wang Y., Zhang Y., Du L., Su G., 2014. Comparison of single-trait and multiple-trait genomic prediction models. BMC Genetics. 15, 30.

Guy D.R., Bishop S.C., Woolliams J.A., Brotherstone S., 2009. Genetic parameters for resistance to Infectious Pancreatic Necrosis in pedigreed Atlantic salmon (Salmo salar) post-smolts using a Reduced Animal Model. Aquaculture. 290, 229-35.

Hayashi T., Iwata H., 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. BMC Bioinformatics. 14, 34.

Hayes B.J., Bowman P.J., Chamberlain A.J., Goddard M.E., 2009. Invited review: Genomic selection in dairy cattle: progress and challenges. Journal of dairy science. 92, 433-43.

Henderson C.R., Quaas R.L., 1976. Multiple Trait Evaluation Using Relatives' Records. Journal of Animal Science. 43, 1188-97.

Houston R.D., Haley C.S., Hamilton A., Guy D.R., Tinch A.E., Taggart J.B., et al., 2008. Major Quantitative Trait Loci Affect Resistance to Infectious Pancreatic Necrosis in Atlantic Salmon (Salmo salar). Genetics. 178, 1109-15.

Jia Y., Jannink J.L., 2012. Multiple-Trait Genomic Selection Methods Increase Genetic Value Prediction Accuracy. Genetics. 192, 1513-22.

Jiang J., Zhang Q., Ma L., Li J., Wang Z., Liu J.F., 2015. Joint prediction of multiple quantitative traits using a Bayesian multivariate antedependence model. Heredity. 115, 29-36.

Lhorente JP G.J., Villanueva B, Carabaño MJ, Neira R, 2014. Coinfection of the Intracellular Bacterial Pathogen Piscirickettsia salmonis and the Sea Louse Caligus rogercresseyi. PLoS ONE 9.

Lillehammer M., Meuwissen T.H.E., Sonesson A.K., 2013. A low-marker density implementation of genomic selection in aquaculture using within-family genomic breeding values. Genetics Selection Evolution. 45, 39.

Luan T., Woolliams J.A., Ødegård J., Dolezal M., Roman-Ponce S.I., Bagnato A., et al., 2012. The importance of identity-by-state information for the accuracy of genomic selection. Genetics Selection Evolution. 44, 28.

Meuwissen T., Goddard M., 2010. The Use of Family Relationships and Linkage Disequilibrium to Impute Phase and Missing Genotypes in Up to Whole-Genome Sequence Density Genotypic Data. Genetics. 185, 1441-9.

Meuwissen T., Hayes B., Goddard M., 2013. Accelerating improvement of livestock with genomic selection. Annual review of animal biosciences. 1, 221-37.

Meuwissen T.H., Hayes B.J., Goddard M.E., 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics. 157, 1819-29.

Meyer K., 2007. WOMBAT – A tool for mixed model analyses in quantitative genetics by REML. J Zhejiang Uni SCIENCE B. 8, 815-21.

Moen T., Baranski M., Sonesson A.K., Kjøglum S., 2009. Confirmation and fine-mapping of a major QTL for resistance to infectious pancreatic necrosis in Atlantic salmon (Salmo salar): population-level associations between markers and trait. BMC Genomics. 10, 368-.

Montesinos-López O.A., Montesinos-López A., Crossa J., Toledo F.H., Pérez-Hernández O., Eskridge K.M., et al., 2016. A Genomic Bayesian Multi-trait and Multi-environment Model. G3: Genes|Genomes|Genetics. 6, 2725-44.

Nielsen H.M., Sonesson A.K., Yazdi H., Meuwissen T.H.E., 2009. Comparison of accuracy of genome-wide and BLUP breeding value estimates in sib based aquaculture breeding schemes. Aquaculture. 289, 259-64.

Nirea K., Sonesson A.K., Lillehammer M., Meuwissen M., 2014.Effect of family structure and size on genomic selection accuracy in aquaculture breeding 10th WCGALP Vancouver, Canada.

Odegard J., Moen T., Santi N., Korsvoll S.A., Kjoglum S., Meuwissen T.H., 2014. Genomic prediction in an admixed population of Atlantic salmon (Salmo salar). Frontiers in genetics. 5, 402.

Pollak E.J., van der Werf J., Quaas R.L., 1984. Selection Bias and Multiple Trait Evaluation. Journal of dairy science. 67, 1590-5.

Schaeffer L.R., 1984. Sire and Cow Evaluation Under Multiple Trait Models. Journal of dairy science. 67, 1567-80.

Solberg T.R., Sonesson A.K., Woolliams J.A., Meuwissen T.H.E., 2008. Genomic selection using different marker types and densities. Journal of Animal Science. 86, 2447-54.

Sonesson A.K., 2007. Within-family marker-assisted selection for aquaculture species. Genetics Selection Evolution. 39, 301.

Sonesson A.K., Meuwissen T.H., 2009. Testing strategies for genomic selection in aquaculture breeding programs. Genetics Selection Evolution. 41, 37.

Sonesson A.K., Meuwissen T.H.E., Goddard M.E., 2010. The use of communal rearing of families and DNA pooling in aquaculture genomic selection schemes. Genetics, Selection, Evolution : GSE. 42, 41.

Ødegård J., Meuwissen T.H., 2014. Identity-by-descent genomic selection using selective and sparse genotyping. Genetics Selection Evolution. 46, 3.

Ødegård J., Meuwissen T.H.E., 2015. Identity-by-descent genomic selection using selective and sparse genotyping for binary traits. Genetics, Selection, Evolution : GSE. 47.