Effect of family structure and size on genomic selection accuracy in aquaculture breeding

K. Nirea¹, A.K. Sonesson² and M. Lillehammer², T. Meuwissen¹.
¹Norwegian University of Life Sciences, ²Nofima AS, Norway

ABSTRACT: A higher accuracy of prediction was obtained for within family genomic selection as compared to the conventional selection method. The accuracy of selection was higher when the family structure was 10x10 followed by 1x10 and 2x2 specifically for within family genomic selection. This was in accordance with the number of full sibs and half sibs produced which increased relationships within or across test and candidate sibs. In all scenarios, accuracy of selection increased as family size increases and the increase was more prominent when family size was beyond 40 individuals per family. Therefore, when resources are limiting, a substantial benefit would be obtained by genotyping 40 to 50 individuals per family. In addition, the benefit would be higher if more related sibs are available in the test and candidate group.

Keywords: Aquaculture; Within-Family Genomic selection; Breeding; Genetics

Introduction

Genomic selection (Meuwissen (2001)) is a novel selection method that increases genetic gain mainly through increasing the prediction accuracy of non-phenotyped individuals. Currently, it is implemented in some terrestrial farm animal species. However, implementation for aquaculture implies genotyping of large numbers of individuals, because breeding in aquaculture species targets both sexes which increases the number of selection candidates to genotype. In addition, estimating breeding value of selection candidates relies on sib-testing which involves a large number of test sibs as compared to other species. Hence, techniques to reduce genotyping costs are needed.

Within family genomic selection was suggested which combines low density molecular and pedigree information as a way of implementing genomic selection for aquaculture breeding programs (Lillehammer et al. (2013); Ødegard and Meuwissen (2014)). Exploiting the full potential of within family genomic selection requires a better understanding of the optimal family structure and the number of individuals per family to genotype which provides an optimal accuracy of selection. Therefore, this study aimed to compare the effects of family structure and family size and their interaction on selection accuracies for within-family genomic selection with conventional BLUP methodology.

Materials and Methods

Genome structure. A population with an effective size of 200 was simulated by randomly selecting and randomly mating 100 sires and 100 dams for 10 000 generations to generate a population in a mutation-drift equilibrium. All individuals had a diploid genome with a length of 100 cM and polymorphisms were created by random mutations. A potential of 3 000 SNPs were selected which have a minor allele frequency (MAF) higher than 0.05 among the SNPs in the course of reaching the Fisher-Wright equilibrium. Among these, a potential of 50 SNP loci were randomly selected to be quantitative trait loci (QTL). The allelic effects of the QTL were independently sampled from a normal distribution. The QTL effects were assumed to be additive and effects were scaled such that the variance and mean of the true breeding value were equal to 1 and 0, respectively.

Breeding schemes: The breeding schemes resembled the structure typical aquaculture breeding schemes. A total of 20 breeding schemes resulting from a combination of 5 family structures and 4 family sizes were simulated. Each breeding schemes had a 100 families and a total population sizes of 2 000, 4 000, 8 000 or 10 000. The number of sires and dams sampled was determined by the type of family structure required. The sirexdam mating ratio were 1x1, 1x2, 2x2, 1x10 or 10x10 resulting in families containing each 20, 40, 80 or 100 individuals. For breeding value estimation, the pedigree was divided into candidate sibs and test sibs where the candidate sibs had 500 individuals and the remaining was set as a test sibs. The candidate sibs were obtained by randomly sampling 5 individuals from each family. All individuals in the pedigree were assumed genotyped and the test sibs were only recorded for a normally distributed single trait with a per chromosome heritability of 0.01.

Breeding value estimation: For the within-family genomic selection schemes, within family breeding values of the candidates were predicted using a linkage analysis based on G matrix, G̃(Luan et al., 2012). The linkage analysis was performed by the LDMP software package (Meuwissen and Goddard (2010)) using two generations of pedigree (the current generation and their parents). Conventional and genomic selection BLUP breeding values were predicted using the ASReml software package (Gilmour et al. (2009)). Both approaches followed similar model differing only in their variance structure:
\[ y = 1\mu + Za + e \]

Where \( y \) is a vector of all phenotypes, \( \mu \) is the overall mean, \( a \) is a vector of additive genetic effects for all animals in the pedigree, \( e \) is a vector of residual random effects and \( Z \) is an incidence matrix linking animals to the records. \( V(a) = A*\sigma^2_a \) for conventional BLUP, and \( V(a) = G_{LA}*\sigma^2_a \) for genomic selection.

Where \( \sigma^2_a \) is the additive genetic variance, \( A \) is a matrix of pedigree relationship and \( G_{LA} \) is a matrix of within-family genomic relationship obtained from the LDMIP software.

**Results and Discussion**

**Effect of family structure on accuracy of selection:** In all scenarios, within family genomic selection led to a substantial increase in accuracy of selection as compared to a traditional selection method (Figure 1). These results were in agreement with previous studies where the classical genomic evaluation method resulted in a higher accuracy of selection and genetic gain as compared with the traditional selection method when similar resources were used (Nielsen et al. (2009)). A higher accuracy of selection was obtained when the family structure was 10x10 followed by 1x10 and 2x2 for both within family genomic selection and traditional selection methods. For conventional selection, the differences in accuracy of selection amongst the family structures were small, except for 10x10 (Figure 1). The increase in accuracy of selection for within family genomic selection was in accordance with the number of full-sibs and half-sibs produced which increases relationships within or across test and candidate groups. This was validated by observing the trend of average relationships generated in the test and candidate groups when different family structures were used (Figure 2). Previous studies have shown similar results where a higher accuracy of selection was observed when a factorial family structure was used for genomic breeding value estimation (Nirea et al. (2012)).

**Effect of family size on accuracy of selection:** Across all scenarios, accuracy of selection increased as family size increased. However, the increase was small when the increase in family size was beyond 40 to 50 individuals per family. Similar results were reported by Lillehammer et al. (2013) where most of potential advantage of within family genomic selection was achieved when 40 individuals per family were genotyped. On the other hand, selective genotyping of 5-20% individuals out of the total available test sibs from both extreme phenotypes of pre-selected families has been shown to result in high accuracy of selection. Therefore, when resources are limiting genotyping could be limited to fewer individuals per family which explain for most of the genetic variability.

**Conclusion**

Results suggest that a significant increase in accuracy of selection could be achieved using more half sibs and/or full sibs in the training and test candidate group. Accuracy of selection also increases when the family size increases. However, the advantage from increasing family size is limited when there are approximately 40 individuals per family, irrespective of mating ratio. Therefore, when resources are limiting, genotyping could be limited in the range of 40 to 50 individuals per family as the increases in accuracy through further increases in family size was moderate. In addition, the benefit would be a higher if more related sibs are available in the test and candidate group.

**Literature Cited**

Genetics. 157: 1819 -1829 
Genet. Sel. Evol. 44: 11 