Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Prioritizing cows for genotyping in Genomic Selection

T. Luan[1], X. J. Yu[1] and T.H.E. Meuwissen[1].
[1]Norwegian University of Life Sciences, Ås, Norway

ABSTRACT: The aim of this research was to investigate the effect of prioritized genotyping cows to improve the accuracy of genomic selection. In the study, TBVs, genotypic and phenotypic data of 326 target bulls, 4,138 training bulls and 5,000 prioritized genotyping cows were simulated based on a real pedigree of dairy cattle. The heritability was 0.8 for bulls and 0.2 for cows. The bulls were 54K genotyped, and cows were 10K genotyped. The GEBVs of target bulls were predicted with training bulls only, and with 1,000, 2,000, 3,000, 4,000 and 5,000 cows included, using GBLUP method. Both weighted and unweighted analyses were carried out. The accuracy was the correlation between GEBVs and TBVs. The results showed that including cows may help to improve the accuracy of the GEBV prediction when reference animals were weighted. When animals were unweighted, including cows didn't improve the accuracy.
Keywords: genomic selection; prioritizing; cow; simulation

Introduction

Genomic selection (GS) is a method to predict breeding values based on the use of dense markers covering the whole genome (Meuwissen et al. (2001)). In GS, first a training population consisting of individuals with both SNP genotypes and phenotypes of a trait is used to construct a model to predict genomic estimated breeding value (GEBV). The model is subsequently applied to the individuals in the test population, which are selection candidates, to predict GEBVs of the individuals based on their SNP genotypes. With the technical advances and development in genotyping technology, the availability of genome-wide, dense molecular markers has enabled GS to become practical in the breeding of several species including dairy cattle breeding. Characterized with long generation and sex-specific phenotype data, dairy cattle breeding can benefit greatly from the implementation of GS (Meuwissen et al. (2001); Schaeffer (2006)).

For a successful implementation of GS, accuracy of the GEBV prediction is a key issue (Goddard and Hayes (2009)). The GEBV prediction accuracy depends on several factors, mainly including (1) linkage disequilibrium (LD) between the markers and quantitative trait locus (QTL); (2) the size of the reference population; (3) heritability of the trait; (4) effective population size Ne; (5) the distribution of QTL effects; (6) relationships between the reference animals and the selection candidates. The factors (2) and (6) may be controlled by breeders, and factor (1) may be improved by a higher marker density. However, the move from a 50k to a 777k SNP panel has not resulted in much improvement, suggesting that the 50k SNP chip is sufficiently dense at least for GBLUP (VanRaden (2013)).

Increasing the size of reference population also helps to improve the accuracy of GEBV prediction (Daetwyler et al. (2008)). In dairy cattle breeding, reference populations have been combined across countries and the reliabilities of genomic prediction was improved relative to using a single-country reference population (Brondum et al. (2011); Lund et al. (2011); VanRaden et al. (2012)). Another way of increasing the size of reference population is to include cows (Zhou et al. (2013)). In dairy cattle GS, usually reference populations comprise progeny-tested bulls to maximize the information from each genotyped individual. However, including cows in the reference population is expected to further increase the information in reference populations and hence improve the accuracy of GEBV prediction.

In dairy cattle breeding, usually the number of cows in the population is much larger than the number of bulls. It is impractical to genotype all the cows due to the economic cost. Therefore rational selection of cows for genotyping is important. To further save cost, cows may be genotyped with a low density SNP chip. The objective of this study was to prioritize cows for genotyping and to evaluate the value of adding low-density genotyped cows to the reference dataset for the GEBV prediction.

Materials and Methods

Simulation. To obtain genotypes and phenotypes of bulls and cows used in the study, a forward simulator was used to simulate populations according to Wright's ideal population model. The effective sizes of the ideal populations were 500, with 1:1 sex ratio. The mutation rate was 10^{-8} per base pair per meiosis. Twenty nine chromosomes of length 1 Morgan were simulated per individual. After 10,000 generations of random mating, the genotypes of the newly produced individuals were recorded. These individuals are referred to as generation 0. The genotypes in generation 0 were gene dropped through a real pedigree of the Norwegian red cattle population. The population consisted of a total of 4,464 genotyped bulls and 815,238 cows in the pedigree. For the bulls, 326 of those born after August, 2012 were set as target bulls. The remaining 4,138 bulls were used as training bulls. For bulls, 1862 SNPs per chromosome were sampled from the genotypes created above. Therefore there were a total of 53,998 markers for 29 chromosomes. For selected cows (described in "Prioritizing" section), 344 SNPs were sampled per chromosome. Both for bulls and cows, 30 QTLs were sampled per chromosome. The minor allele frequency for markers and QTLs was 0.05. Additive genetic effects were determined by 870 QTLs. QTL effects were drawn from a Laplacian distribution with mean 0 and shape parameter 1. It was assumed that all QTL effects were additive. True breeding values (TBV) were calculated by summing all QTL effects. The phenotype was simulated as the sum of TBV and a random environmental effect in order to achieve a heritability of daughter-yield-deviations (DYDs) of bulls of 0.80 and 0.20 for phenotypes of cows. The phenotypes of target bulls were set unknown and predicted using the method described in "Statistical analysis" section. In the study, both weighted (with reliability of phenotype) and unweighted analyses were carried out. In the unweighted analysis, DYDs of training bulls and phenotypes of cows were treated equally in the model described in "Statistical analysis" section to predict GEBV for target bulls. In the weighted analysis, DYDs were weighted using a weighting factor 16 since a DYD with heritability 0.8 for training bulls has the same information content as 16 repeated records with heritability 0.2 for cows (Garrick et. al (2009)).

Prioritizing. The cows were prioritized by minimizing the conditional variance of the target bull population criterion (Yu et al., 2014; submitted), where the genetic variance of the target bulls, represented by their pedigree relationship matrix A, was conditioned on the cow suggested for genotyping, i.e.:

$$A_{ii}^{*} = A_{ii} - h^{2}A_{ik}A_{jk}/A_{kk}$$

where A_{ii}^{*} is the conditional relationship of a target bull, which is to be minimized, and k is a cow considered for prioritization; and h^2 accounts for the trait heritability. The cow k that minimizes the average A_{ii}^{*} of the target bulls, is selected for genotyping. Next the relationship matrix of all animals is conditioned on this cow using the above formula, and the next cow k' (≠k) is selected given that she further reduces the conditional relationship of the target bulls. This procedure is repeated until the desired number of cows is prioritized.

Statistical analyses. GEBV of 326 target bulls were predicted with the GBLUP model. GBLUP estimates the effects of the markers by best linear unbiased prediction (Meuwissen et al. (2001)), assuming that every marker explains an equal proportion of the total genetic variance. The model was

$$y = \mu + \sum_{j=1}^{N_m} X_j a_j + e$$

where y is a $N \times 1$ vector of phenotypes, $\boldsymbol{N_m}$ is the number of markers fitted; a_j is the effect of the marker; X_j is a $N \times 1$ vector denoting the genotype of the individuals for marker j, with $X_{ij} = -2q_j/\sqrt{H_j}$ if individual i is homozygous for the first allele at locus j, $X_{ij} = (1-2q_j)\sqrt{H_j}$ if individual i is heterozygous, $X_{ij} = (2-2q_j)/\sqrt{H_j}$ if individual i is homozygous for the second allele at locus j, and $X_{ij} = 0$ if the marker genotype is missing, where q_j is the frequency of the second marker allele and H_j is the marker heterozygosity. The division by $\sqrt{H_j}$ standardizes the variance of the marker genotype data to 1. Given the estimates of the marker effects and the marker genotypes, the genomic breeding values of the target bulls are predicted as $GEBV_i = \sum_{j=1}^{N_m} X_{ij}\hat{a}_j$ where X_{ij} is the marker genotype of individual i for marker j coded the same as above, and $\hat{a}_j$ is the estimated effect of marker j.

Results and Discussion

Table 1 shows the accuracy of GEBV prediction using only training bulls, and including 1,000, 2,000, 3,000, 4,000 and 5,000 cows. Phenotypes of reference populations were weighted with a factor 16 or equally weighted. The results showed a different trend of accuracy of the predictions with or without weight. For GEBV prediction with weighted data, the accuracy of the prediction including cows was higher than the prediction using only the training bulls. The accuracy increased when including up to 3,000 cows, with very little improvement when more cows were included in the reference population. It is expected that including cows in the reference population may increase the information to be used for prediction and therefore the accuracy increases. A weighting factor of 16 implies that the phenotype of training bulls was measured 16 times while that of cows was measured once. Thus the contribution of the high quality information of the bulls was emphasized, while information of low accuracy phenotypes of cows was still taken into account to predict the GEBV.

Table 1. Accuracy of GEBV prediction including different number of prioritized genotyping cows.

	weight (bull:cow)	
Number	16:1	1:1
0	0.7271	0.7216
1000	0.7302	0.7073
2000	0.7320	0.7010
3000	0.7334	0.7035
4000	0.7339	0.6991
5000	0.7335	0.6947

For the prediction with unweighted data, the accuracy of the prediction using only the training bulls was the highest, indicating including cows did not help to improve the prediction. For example, a ~2% decrease of accuracy was observed when adding 1,000 cows in the reference data set compared to using only training bulls. In addition, the results showed that the accuracy decreased with the increase of the cows in the reference population. The cows were treated equally important as bulls in unweighted reference data set, which may lead to the reference data including cows yielded poorer information to predict GEBV compared to the reference data that only includes training bulls, and thus deteriorated the accuracy of the prediction.

Conclusion

In the presented study we evaluated the GEBV prediction when prioritizing cows for inclusion in the reference data set. Results suggest that when animals were weighted differentially based on the accuracy of their phenotypic data, including cows in the reference data can help to improve the accuracy of the prediction somewhat. However, when animals were unweighted, including cows in the reference data did not improve the accuracy. Hence, the inclusion of cows the in the reference population should be accompanied by an accurate weighting of the phenotypic information on bulls versus that on cows.

Literature Cited

Brondum, R.F., Rius-Vilarrasa, E., Stranden, I. et al. (2011). J. Dairy. Sci. 94:4700-4707.

Daetwyler, H.D., Villanueva, B., and Woolliams, J.A. (2008). PLoS One 3:e3395.

Garrick, D.J., Taylor, J.F., and Fernando, R.L. (2009). Genet. Sel. Evol. 41:55.

Goddard, M. E., and Hayes, B. J. (2009). Nat. Rev. Genet. 10: 381-391.

Lund, M.S., de Roos, A.P.W., de Vries, A.G. et al. (2011). Genet. Sel. Evol. 43:43

Meuwissen, T.H.E., Hayes, B.J., and Goddard, M.E. (2001). Genetics 157: 1819-1829.

Schaeffer, L. R. (2006). J. Anim. Breed. Genet. 123: 218-223.

VanRaden, P.M., Olson, K.M., Null, D.J. et al. (2012). Interbull Bulletin No. 46:75-79.

VanRaden, P.M., Null, D.J., Sargolzael, M. et al. (2013). J. Dairy Sci. 96:668-678.

Yu, X. J., Woolliams, J.A., and Meuwissen, T.H.E. (2014). Genet. Sel. Evol. (submitted).

Zhou, L., Ding , X., Zhang, Q., et al. (2013). Genet. Sel. Evol. 45:7