The impact of using individuals from different generations with different accuracy of BLUP EBVs on accuracy of Genomic EBVs

M. Saatchi[*], A. Nejati-Javaremi, S. R. Miraei-Ashtiani, H. Mehrabani-Yeganeh, M. Moradi-Shahrebabak

Introduction

It is currently possible to genotype individuals for 30000, 50000 or more single nucleotide polymorphisms (SNP) with gene chip array, and the bovine genome can be covered with a dense SNP map. These markers can be used to estimate genome-assisted breeding values (GEBVs) as proposed by Meuwissen et al. (2001). Application of genomic selection may considerably decrease the costs of animal breeding program up to 90% (Schaeffer, 2006). Individuals with phenotypic records may belong to different generations. For example, in dairy cattle, bulls' DNA exists from several generations. Saatchi et al. (2010) has shown that using individuals of more recent generations in training set led to higher accuracy of genomic estimated breeding values (GEBVs) than individuals from more distant generations.

In dairy cattle; bulls don't have phenotype records for most economic traits. But their breeding values are mostly estimated by best linear unbiased prediction (BLUP) methods. Old bulls usually have highly accurate BLUP EBVs while recent bulls have lower BLUP EBVs accuracy. Saatchi et al. (unpublished) showed that whatever accuracy of BLUP EBVs is increase then accuracy of GEBVs will increase. The main question is that which ones (old bulls with highly accurate BLUP EBVs versus recent bulls with lower BLUP EBVs accuracy) leads to higher accuracy of GEBVs? Different scenarios were compared by including individuals with different accuracy of BLUP EBV's from different generations.

Materials and methods

Simulation:

A genome consist one chromosome with 100 cM length and 1000 equally spaced SNPs (every 0.1 cM) and 10 QTLs was generated for each individual. This genome length was not realistic, but the density of the markers is comparable to a 30000 SNPs on a 30-M genome. Both SNP and QTL were assumed to be biallelic with equal initial allele frequency. For these simulations gene substitution effects for each QTL loci were assigned randomly from a normal distribution with a mean zero and unit variance, $a \sim N(0, 1)$. Ten QTLs covered all the genetic variance and animal true breeding values was then sum of the 10 QTLs effects. Only additive genetic effects were considered.

An effective population size of 100 animals was simulated, of which 50 animals were male and 50 female. This structure followed by 50 generations of random mating, implying that each animal had on average two offspring in the next generation. The paternal and maternal

[*] Department of Animal Science, University College of Agriculture & Natural Resources, University of Tehran, Karaj, Iran.

haplotypes for each animal were generated based on each parent using the Haldane mapping function for generation of recombinant haplotypes. Sires and dams in the base generation were assumed to be unrelated. Fifty generation of random mating were made to generate sufficient linkage disequilibrium (LD) between loci.
After the first 50 generations, 9 additional generations (51 to 59) were simulated. In the 51nd generation, the population was expanded to obtain a population size of 500 males and 500 females. In the following generations, 500 sires are mated to 500 dams randomly in each generation, which are discrete. Markers data were recorded for sires of last nine generation (51-59), but only sires of generation 51-58 had trait phenotype (BLUP EBVs) and thus were included in the training set according to different scenarios. The validation set contained 500 sires from generation 59.
Breeding values with different accuracy for sires were simulated to allow comparisons, in terms of accuracy of predicted genomic breeding values. For simulation of estimated breeding value (BLUP EBVs) of sires according to required accuracy, we used below equation (Nejati-Javaremi, unpublished):

$$EBV = TBV \times Accuracy + Z_i \times \sqrt{(1 - Accuracy^2) \times S^2_a}$$

Where, TBV is sire' true breeding value and $Z_i \sim N(0, 1)$.

Models

For calculation of marker effects, the simple mixed model estimation method suggested by Meuwissen et al. (2001) was used assuming that all loci explained a priori an equal amount of variance. The genetic variance among individuals was determined as variance of true breeding value of males in generation 51-58. The model to estimate the marker effects was

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Zm} + \mathbf{e} \quad (1)$$

Where, **y** is sires' BLUP EBVs, **b** is vector of mean, **m** is vector of random marker effects, **e** is vector of random residual error, **X** and **Z** is coefficient matrix. Row elements of **Z** consist of 0, 1 and 2 for marker genotype.
After solving equations and estimation of marker effects, total merit or GEBV, was estimated as

$$\mathbf{GEBV_i} = \mathbf{Z_i m_i} \quad (2)$$

Different scenarios were compared by including individuals with different accuracy of BLUP EBV's from different generations in training set. Accuracies were calculated as the correlation between simulated and estimated breeding values. Each simulated data set and model analysis was replicated 10 times and results were averaged across replicates.

Results

Results from using sires from different generations with different accuracy of BLUP EBVs in training set, on accuracy of GEBVs are presented in Table 1.

Table 1- Accuracy of GEBVs in training set and validation set using sires from different generations with different accuracy of BLUP EBVs in training set.

sires used in training set (their BLUP EBVs accuracy)	Accuracy of GEBVs in training set	Accuracy of GEBVs in validation set
Generation 51 (0.99)	0.988± 0.001	0.914± 0.007
Generation 52 (0.99)	0.986± 0.001	0.924± 0.006
Generation 53 (0.99)	0.983± 0.001	0.933± 0.006
Generation 54 (0.99)	0.982± 0.001	0.936± 0.006
Generation 55 (0.99)	0.982± 0.001	0.941± 0.005
Generation 56 (0.99)	0.980± 0.001	0.946± 0.005
Generation 57 (0.99)	0.980± 0.001	0.949± 0.005
Generation 58 (0.99)	0.980± 0.001	0.953± 0.004
Generation 58 (0.90)	0.952± 0.003	0.922± 0.006
Generation 58 (0.80)	0.919± 0.003	0.880± 0.008
Generation 58 (0.70)	0.889± 0.004	0.845± 0.010
Generation 58 (0.50)	0.816± 0.010	0.768± 0.013

As it is shown in Table 1, in condition of same BLUP EBVs accuracy of training set individuals (accuracy equal to 0.99) and same number of individuals in training set (500 sires), when individuals in training set belong to generations close to validation set (generation 59), accuracy of GEBV is higher than when they belong to generations far from validation set. This result is in agreement with previous studies (Nejati-Javaremi et al., 1997; Saatchi et al., 2009). This may be due to: (i) weaker relationship between individuals of training set and validation set, (ii) higher amounts of recombination and changes in haplotypes structure by passing generations.

In condition of different BLUP EBVs accuracy of recent generation (generation 58), results show that by decreasing in accuracy of BLUP EBVs, accuracy of GEBVs in training set and validation set will decrease.

Comparison results of using individuals from recent generation (generation 58) with different accuracy of BLUP EBVs in training set with individuals from older generations with highly accurate BLUP EBVs, show that accuracy of GEBVs obtained by using individuals with BLUP EBVs accuracy equal 0.9 from recent generation in training set is the same with using individuals with highly accurate BLUP EBVs (0.99) from older generations. When accuracy of BLUP EBVs of recent generation individuals are lower than 0.90, accuracy of obtained GEBVs is lower than using individuals with highly accurate BLUP EBVs even from very old generations. This results show that the impact of accuracy of BLUP EBVs as phenotype records on accuracy of GEBVs are more important than the impact of generation distance between training set and validation set. The same results are reported in other studies when phenotype records of training set individuals are used for estimation of marker effects instead of used BLUP EBVs in this study. Villumsen et al. (2009) showed that the measure of reduction in reliability of GEBVs in next seven generations after estimation of marker effects is low and is about 1 to 3 percent per generation. While, in their study showed that the impact of heritability of trait on reliability of GEBVs is high in different scenarios. Similar results are obtained in Muir (2007) study.

Conclusion

According to these results, , it could be recommended to use individuals from recent generation with highly accurate BLUP EBVs in training set. Proved bulls from recent generation with first crop daughters usually have BLUP EBVs with accuracy above 0.9. Then, it is safe to comment that these bulls could be used in training set besides bulls from older generation with high accurate BLUP EBVs.

References

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

Muir, W. M. (2007). *J Anim. Breed. Genet*. 124:342-355.

Nejati-Javaremi A., Smith C., Gibson J. P. (1997). *J. Anim. Sci*., 75, 1738-1745.

Saatchi M., Miraei-Ashtiani S. R., Nejati-Javaremi A., *et al*. (2010). *African J. of Biotechnology,* 9(4), 438-442.

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

Villumsen T. M., Janss L., Lund M. S. (2009*). J. Anim. Breed. Genet*., 126, 3-13.