Available online at www.scholarsresearchlibrary.com

Scholars Research Library

Scholars Research Library

Annals of Biological Research, 2013, 4 (9):92-99
(http://scholarsresearchlibrary.com/archive.html)

Annals of Biological Research
Scholars Research Library
ISSN 0976-1233
CODEN (USA): ABRNBW

The effect of genomic selection and classic method on accuracy of breeding value in threshold traits

Reza Behmaram[1*], Ali Asghar Aslaminejad[1], Mehdi Aminafshar[2] and Mojtaba Tahmoorespur[1]

[1]Department of Animal Science, Faculty of Agriculture, Ferdowsi University of Mashhad, Mashhad, Iran
[2]Department of Animal Science, Faculty of Agriculture and Natural Resources, Science and Research Branch, Islamic Azad University, Tehran, Iran

---

ABSTRACT

With the availability of high-density marker maps and cost-effective genotyping, genomic selection(GS) methods may provide faster genetic gain than can be achieved by current selection methods based on phenotypes and the pedigree. Many schemes have been proposed for continuous traits, but methods for threshold traits are still scarce. Accuracies for breeding values were investigated for a typical dairy cattle breeding setting by using genomic and classic methods. Here we investigate some of the factors driving the accuracy of genomic selection, namely marker numbers and heritability. In classic method, we estimated true breeding value(TBV) using ASReml from univariate analysis model for threshold traits. Marker characteristics and linkage disequilibrium were obtained by simulation to achieve a mutation drift balance. Six generations with only genotypes were generated to examine accuracy changes rate over time. With h^2= 0.05, accuracies for genomic and classic path ranged from 0.22 to 0.45 and 0.15 to .035 respectively. For genomic and classic methods with h^2= 0.30, accuracies varied from 0.27 to 0.61 and 0.22 to 0.44 respectively. With h^2= 0.80, accuracies for genomic and classic approaches ranged from 0.21 to 0.73 and 0.36 to .055 respectively. Results showed accuracies of breeding value by genomic selection were sufficiently high to implement dairy selection schemes testing in which case a data time-lag of two to three generations may be present. Using traditional method for estimating TBV resulted in reduced accuracies compared with direct genomic selection.

Keywords: accuracy, genomic selection, heritability, marker ,threshold traits.

INTRODUCTION

To overcome the problem with traditional marker-assisted selection (MAS) that only covers a limited proportion of total genetic variance by the markers, a new technique called genomic selection was presented by Meuwissen et al. [16] ,which traces all quantitative trait loci(QTL) by tracing all chromosome segments through highly dense markers covering the entire genome.

GS has became feasible very recently with the availability of high-through put genotyping technology. Estimation of genomic breeding value(GEBV) is the key step in GS, for many of approaches have been proposed [2, 3, 5, 15, 16]. To estimate GEBV, a prediction equation based on the single nucleotide polymorphism(SNP) is first derived. The whole genome is divided into small segments, the effects of which are estimated in a reference population in which animals are both phenotyped and genotyped. Then, the effects of all loci that contribute to genetic variation are captured, even if the effects of the individuals loci are very small. In subsequent generations, animals can be genotyped for the markers to determine which chromosome segments they carry, and the estimated effects of the segments the animal carries can then be summed across the whole genome to predict the GEBV. This breeding value is termed a GEBV.

Meuwissen et al. [16] demonstrated in simulations that it was possible to achieve accuracies of predicted breeding values from markers alone of 0.85 (where accuracy is the correlation between true breeding value (TBV) and estimated breeding value (EBV), and the reliability is the square of this result).

Most of the estimation methods focus on continuous traits. However, many traits of importance in animal production, such as litter size of large mammals, degree of calving difficulty and resistance to disease, present a discrete(categorical) distribution of phenotypes, and are often termed threshold traits. Obviously, the GS methods proposed for continuous traits cannot be adequately applied for such kind of traits.

This research was performed to investigate the accuracy of breeding values from genomic selection for a threshold trait is higher than classical method or no and the amount of their difference.

MATERIALS AND METHODS

Simulation

A genome consisting 3 chromosomes each with 100 cM in length with 100, 200, 400, 800 equally spaced single nucleotide polymorphism(SNP) (each 1 cM) and a total number of 30, 60, 120, 240 QTLs (that scattered on chromosomes randomly) was generated for each individual. This small genome size was chosen to decrease the calculation time.

Both SNP and QTL were assumed to be biallelic with equal initial allelic frequencies. For these simulations, gene substitution effects for each QTL were assigned randomly from a standard normal distribution, a ~ N (0, 1). QTLs covered total genetic variance and individual true breeding values. Only additive genetic effect was considered.

An effective population size of 100 individuals was simulated, of which 50 were male and 50 were female. This structure was followed by 50 generations of random mating, implying that each individual had on average two offspring in the next generation (variance of family size was two).

The paternal and maternal haplotypes for each individual were generated based on Haldane mapping function to generate recombinant haplotypes. Sires and dams in the base generation were assumed to be unrelated.

Fifty generations of random mating were practiced to generate sufficient linkage disequilibrium (LD) between loci. Two LD measurements, r^2 and D' , were used to calculate LD in generation 50, as average of all synthetic marker loci. Markers with a minor allele frequency of < 0.05 were discarded. After the first 50 generations, 6 additional generations (51 to 56) were simulated. Population was expanded to obtain intended population size in generation 51. Population size was constant until generation 56. For population size, 1000 individuals with equal number of males and females in each of the last 6 generations were simulated. Only females of generations 51 through 55 (500 females in each generation) had trait phenotype and, thus, were included in the training set according to different scenarios.

To investigate the effect of generation distance between training set and validation set on accuracy of GEBVs, females from different generations (distant and recent generations) were included in training set.

The validation data contained individuals from generation 56. For simplification, no selection was considered. Population structure and parameters used in the simulation are presented in Table 1.

Analysis model

For calculation of GEBV, the simple mixed model estimation method suggested by Meuwissen et al. [16] was used assuming that all loci explained and equal amount of variance (That is, the variance per locus σ^2m, is σ^2m = σ^2a /n where σ^2a is the total genetic variance and *n* is the number of marker loci).

Meuwissen et al. [16] termed this method as best linear unbiased prediction (BLUP). This assumption (equal variance over all loci) is clearly unrealistic. Genetic variance may not be equal across markers, for example, major genes may exist on some chromosomes. However, BLUP is quick, easy to program and as Meuwissen et al. [16] demonstrated, BLUP performs almost as well as the much more advanced and time consuming Bayesian methods.

The model to estimate the marker effects was

$$y = Xb + Zm + e \qquad (1)$$

where, y is the vector of observations, b is the vector of means, m is the vector of random marker effects, e is the vector of random residual effects, X and Z are coefficient matrices. Row elements of Z consist of 0, 1 and 2 for marker genotype.

Then, the expected value of y is 1μ and the variance of y is

$$V(y) = ZIZ'\sigma^2 m + I\sigma^2 e \qquad (2)$$

(assuming equal variance for each marker).

The mixed model equation (MME) for BLUP is

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\alpha \end{bmatrix} \begin{bmatrix} b \\ m \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix} \qquad (3)$$

We considered $\alpha = \sigma^2 e / \sigma^2 m$ as Meuwissen et al. [16]. After obtaining solution for vector m, GEBV was estimated as

$$GEBV_i = Z_i m_i \qquad (4)$$

The genetic variance was determined as variance of true breeding values among individuals in generation 51 through 55. As haplotyping would increase computation time with little or no gain in accuracy at high marker density [13].

We used genotypes rather than haplotypes Each simulated data set was replicated 10 times and results were averaged across replicates.

True breeding value

We estimated TBV from two path. Once from the above stages, we calculated it simultaneously by GEBV from genomic path. For estimating TBV from classic path, based on the suitable single- trait animal model for threshold traits using ASREML, including the mean as fixed effect in model, breeding values were estimated.
Then we put this breeding value as TBV in accuracy formula and calculated its correlation with GEBV as accuracy. (accuracies were calculated as the correlation between simulated and classic TBV and GEBV).

Table 1. Population structure and parameters used in the simulation.

Parameter	Value
Number of chromosome	3
Number of SNP markers per chromosome	100, 200, 400,800
Genome Length (cM)	300
Marker distance (cM)	1
Number of QTL	30, 60, 120, 240
QTL effects	Normal distribution
Recombination	Haldane map function
Number of generation	56
Generation 1 to 50, create LD	50 male, 50 female
Generation 51 to 56	500 male, 500 female
Training set	Females of generation 51 to 55
Validation set	Females of generation 56
Heritability	0.05, 0.30, 0.80

Table 2. Mean (±SE) of homozygosity and linkage disequilibrium (D′ and r^2) between markers in generation 50.

Parameter	Mean ± SE
D′	0.61 ± 0.003
r^2	0.18 ± 0.002
Homozygosity	0.58 ± 0.002

Table 3. Number of Markers and QTLs in classic method.

Scenarios	Details
A	100 markers, 30 QTL
B	200 markers, 60 QTL
C	400 markers, 120 QTL
D	800 markers, 240 QTL

RESULTS AND DISCUSSION

Results

The presented simulations assumed a small effective population size of Ne = 100, which generates LD between the markers and a QTL and thus causes the marker effects. The expected amount of disequilibrium in a stable population represents a balance between its creation by drift and its decay by recombination. In this research, six generations without own performance with three different levels of heritability (0.05, 0.30, 0.80) and different numbers of markers (100, 200, 400, 800) were simulated simultaneously.

We examined the effect of calculating true breeding from traditional method on accuracy rate and comparing it with genomic path in different schemes. All the accuracies of selection are given in tables 4 to 8. Tables 4 to 7 showed the correlations between GEBVs and simulated TBV and table 8 presented correlation between GEBVs and calculating TBV from traditional method.

The highest accuracy for genomic and classic path was 0.73 and 0.55 respectively. The lowest accuracy for them was 0.21 and 0.15 respectively. The results showed a relatively clear relationship between the number of markers used in the prediction model and the accuracies that were obtained.

The accuracy in genomic selection varied from 0.21 to 0.64 and 0.24 to 0.70 using 100 and 200 markers respectively . Whereas, estimating TBV from classic method resulted in accuracies ranged from 0.15 to 0.36 and 0.23 to 043 for 100 and 200 markers respectively.

The accuracy ranged from 0.26 to 0.73 and 0.28 to 0.68 using 400 and 800 markers respectively in genomic path. On the other hand, with estimating TBV by traditional approach, accuracies varied from 0.29 to 0.46 and 0.35 to 0.55 for 400 and 800 markers respectively.

The regression of TBV on EBV become closer to 1 as the marker density increased. In general, the density of the markers increased the accuracy of estimating the breeding values as expected. Tables 4 to 7 showed higher heritability results in higher accuracies for all the generations in genomic selection.

When markers number increased from 400 to 800 we saw a partial decay for accuracies, although we did not expected it. When we used traditional method for estimating TBV, with heritability 0.05, 0.30, 0.80 accuracies ranged from 0.15 to 0.35, 0.22 to 0.44 and from 0.36 to 0.55.

On the other hand, for genomic path, with a heritability of (0.05) the accuracy in generation 51 was 0.45 decreasing to 0.22 in generation 56. Almost the same patterns were observed for the other values. With a heritability of (0.30) the accuracy in generation 51 was 0.61 decreasing to 0.27 in generation 56. With a heritability of (0.80) the accuracy in generation 51 was 0.73 decreasing to 0.21 in generation 56.

Table 4. Accuracies of breeding value in generation 51 to 56 (Number of Markers= 100, Number of QTL = 30) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.43	0.27	0.27	0.26	0.24	0.22
$h^2 = 0.30$	0.60	0.36	0.34	0.31	0.29	0.27
$h^2 = 0.80$	0.64	0.35	0.33	0.26	0.24	0.21

Table5. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 200, Number of QTL = 60) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.42	0.28	0.32	0.28	0.25	0.24
$h^2 = 0.30$	0.61	0.41	0.42	0.38	0.35	0.34
$h^2 = 0.80$	0.70	0.42	0.40	0.34	0.33	0.29

There is a tendency for the average decay to be higher from generation 51 to 52 than from generation 52 to 56. Our findings indicate that higher heritability results in almost higher accuracy for all the generations 51 to 56.

On the other hand, when we estimated TBV from classic method by ASREML, and then, calculated its correlation with GEBV as accuracy of breeding value, we found lower values for it rather than when we estimated TBV by genomic selection in all of scenarios.

Table 6. Accuracies of breeding value in generation 51 to 56 (Number of Markers= 400, Number of QTL = 120) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.45	0.32	0.31	0.30	0.27	0.26
$h^2 = 0.30$	0.59	0.42	0.42	0.41	0.37	0.34
$h^2 = 0.80$	0.73	0.46	0.44	0.41	0.38	0.36

Table 7. Accuracies of breeding value in generation 51 to 56 (Number of Markers = 800, Number of QTL = 240) for threshold traits.

Generation	51	52	53	54	55	56
$h^2 = 0.05$	0.44	0.29	0.31	0.31	0.30	0.28
$h^2 = 0.30$	0.59	0.44	0.42	0.41	0.41	0.40
$h^2 = 0.80$	0.68	0.42	0.41	0.38	0.37	0.36

Table 8. Accuracies of breeding value for threshold traits using by classical method

Scenario	A	B	C	D
$h^2 = 0.05$	0.15	0.23	0.29	0.35
$h^2 = 0.30$	0.22	0.30	0.37	0.44
$h^2 = 0.80$	0.36	0.43	0.46	0.55

Discussion

The estimation model assumes that there is no dominance (i.e., only the additive effects are fitted), and the average effects of the genes are estimated, which is probably satisfactory for the prediction of breeding values in most cases. It is important to notice that in this study, the population is random mating. Muir [18] reported the decay of accuracies were faster in situations with directional selection compared with random mating due to changes in allele frequency, genetic variance and LD in each generation.

The accuracies of selection are given in tables 4 to 8. In all scenarios we used classic TBV, lower accuracy was observed comparing with similar conditions at genomic path. In a similar research for traditional scheme in Canadian Holsteins, the accuracy of predicting the EBV of progeny-tested young bulls was estimated to be 0.75 for their first EBV [10].

A relationship between heritability and accuracies was observed, as heritability decreased so did the accuracy. Meuwissen et al. [16] showed that the accuracy of GEBVs decreased to 0.804, 0.768, 0.758, 0.734 and 0.718 in 5 subsequent generations, respectively. This phenomena was also reported by Kolbehdari et al. [9].

In other research was done by Wang et al. [4] about threshold traits, the accuracies declined over generations for all methods with almost the same rate and by decreasing the heritability from 0.50 to 0.05, the accuracies decreased.

Some studies suggested it can be compensated by using a larger number of observations like Calus & Veerkamp [12].

On the other hand, Habier et al. [8] indicated that using genomic relationship among individuals and LD between markers and QTL result in higher accuracy because of including genomic relationship information among individuals. In an earlier study, Nejati- Javaremi et al. [1] documented the impact of replacing pedigree by marker on higher accuracy of evaluation.

GS can be useful for traits with low heritability by using more balanced selection scheme and appropriate models. With traditional selection schemes, low heritable traits are only improved slowly because reliability of the breeding value depends strongly on the heritability. With GS the reliability depends on many factors like size of dataset used to make the marker associations, recombinations, LD and marker density which is similar for both low and high heritability traits.

Therefore, use of GS may lead to a more balanced selection index than obtained in the traditional selection schemes for two reasons. First, GS models seem to perform well for low heritable traits, where the reliabilities could be as high as or higher than obtained in a traditional selection scheme [10]. Second, important low heritable traits like disease traits are often not included in the breeding objective because getting accurate records on these traits is difficult. With GS the having phenotypes on close relatives of breeding candidates is not necessary, as the prediction ability persists over several generations. This gives the opportunity to perform difficult recordings in designated herds, which are then used to construct prediction models for the entire population.

Advantages of Bayesian method to BLUP evaluation has been shown in some studies [2, 16]. For example, Meuwissen et al. [16] used Bayesian method and obtained accuracy of 0.848 for the GEBVs of individuals in the training set.

Because of the only gradual decay in the accuracy of selection over time, it can be concluded that GS will be well applicable to dairy breeding setting without progeny testing. Therefore, costs of breeding may be reduced and in addition, genetic gain per year is accelerated by reducing the generation interval.

Implementation of GS for dairy breeding requires high reliabilities for GEBV for at least two to three generations ahead without having phenotypes. GS is therefore possible and a very interesting approach to replace or supplement progeny testing.

On the other hand, we observed a tendency for the rate of decay to be higher for the first two to three generations, which is the critical time frame for the use of genomic predictions in dairy cattle. It could be interesting to understand the causes of this decay and to see if relative improvements in the genomic prediction models could be made to keep reliabilities high over a longer time span.

In genomic selection, effects of QTL are distributed among adjacent marker loci. In other words, some degrees of co-linearity exist among neighboring markers. With increasing distance between generations in training set and generation of validation set, because of higher amounts of recombination occurrence the accuracy of evaluation decreases.

Muir [18] suggested that after several generations following estimation of marker effect, the accuracy reduces and these effects should be re estimated. Similar results have been reported elsewhere [11, 17]. The basis for the high reduction and possible improvements should be investigated more thoroughly.

In practical situations, this means that it is advantageous to use genomic method because this avoids the traditional estimation of TBV with the associated problems such as unavailable or incomplete recording data.

However, if cost of genotyping is an issue it may be recommended to use genotypes and phenotypic information of individuals. GS has revolutionized dairy cattle breeding by greatly increasing the accuracies of estimates of genetic merit for young animals and could double the rate of genetic progress by shortening the generation interval. GS so far has focused on continuous traits, although, many threshold traits significantly affect profitability and are difficult to be selected.

Therefore, GS for threshold traits is important in animal breeding. Among many existing approaches for estimating genomic breeding values of quantitative traits, the three normal Bayesian methods (BayesA, BayesB and BayesCp) are commonly used. But they are not suitable for threshold traits, because they are based on linear models.

CONCLUSION

Genomic selection can be used for threshold traits like continuous traits in practical animal breeding because of having many advantages compare with traditional methods, although, statistical methods and analyze models should be investigated more.

REFERENCES

[1] A. Nejati-Javaremi, C. Smith, J. P. Gibson, *J. Anim. Sci*, **1997**, 75, 1738-1745.
[2] B. J. Hayes, P. J. Bowman, A. J. Chamberlin, M. E. Goddard, *J. Dairy Sci*, **2009**, 92, 433-443.
[3] B.J. Hayes, P.J. Bowman, A. C. Chamberlain, K. Verbyla, M. E. Goddard, *Genet. Sel. Evol*, **2009**, 41, 41-51.
[4] C. L. Wang, X.D. Ding, J.Y. Wang, J.F. Liu, W.X. Fu, Z.J. Yin , Q. Zhang, *Heredity*, **2013**, 110, 213-219.
[5] D. Gianola, R. L. Fernando, A. Stella. *Genetics*, **2006**,173, 1761-1776.
[6] D. Habier, J. Tetens, F. R. Seefried, P. Lichtner, G. Thaller, *Genet. Sel. Evol*, **2010**, 42, 5-17.
[7] D. Habier, R. L. Fernando, K. Kizilkaya, D. J. Garrick, BMB *Bioinformatics*, **2011**, 186-194.
[8] D. Habier, R. L. Fernando, J. C. M. Dekkers, *Genetics*, **2007**, 177, 2389-2397.
[9] D. Kolbehdari, L. R. Schaeffer, J. A. B. Robinson, *J. Anim. Breed. Genet*, **2007**, 124, 356-361.
[10] L. R. Schaeffer, *J. Anim. Breed. Genet*, **2006**, 123, 218-223.
[11] M. E. Goddard, *Genetica*, **2009**, 136, 245-257.
[12] M. P. L. Calus, R. F. Veerkamp. **2007**, *J. Anim. Breed. Genet*, 124, 362-368.
[13] M. P. L. Calus, T. H. E. Meuwissen, A. P. W. Deroos, R. F. Veerkamp, *Genetics*, **2008**, 178, 553-561
[14] P. M. VanRadan, J. Dairy Sci, **2008**, 91, 4414-4423.
[15] T. R. Solberg, A. K. Sonesson, J. A. Woolliams, T. H. E. Meuwissen, *J. Anim. Sci*, **2008**, 86, 2447-2454.
[16] T. H. E. Meuwissen, B. J. Hayes, M. E. Goddard, *Genetics*, **2001**, 157, 1819-1829.
[17] T. M. Villumsen, L. Janss, M. S. Lund, *J. Anim. Breed. Genet*, **2009**, 126, 3-13.
[18] W. M. Muir, *J Anim. Breed. Genet*, **2007**, 124, 342-355.