African Journal of Biotechnology Vol. 10(33), pp. 6198-6202, 6 July, 2011
Available online at http://www.academicjournals.org/AJB
DOI: 10.5897/AJB10.1025
ISSN 1684–5315 © 2011 Academic Journals

Full Length Research Paper

Multiple-trait genetic evaluation using genomic matrix

Mahboubeh Rostami-Angasi* and Ahmad Sodagar-Amiri

Chalous Branch, Islamic Azad University, Chalous, Iran.

Accepted 17 March, 2011

Accuracy of the multiple-trait genetic evaluation based on genomic matrix obtained from allelic relationships was estimated through computer simulation and was compared with the accuracy of traditional Best Linear Unbiased Prediction (BLUP). Firstly, a base population (Ne = 100) was simulated and for each animal in the base population, three chromosomes were created. On each chromosome, 200 markers and 50 quantitative trait loci (QTLs) were randomly located. After 110 generations of random mating, linkage disequilibrium was created between the marker and QTL. Multiple-trait evaluation was done for two traits with high (h^2 = 0.46) and low heritability (h^2 = 0.1). In the first trial, in order to study the changes of evaluation accuracy along generations, after creating linkage disequilibrium, the population size of the last generation was expanded 3 times and random mating was done for the next three generations. Then, phenotypic and genotypic records of females for the last three generations were simulated. The results showed that the accuracy of evaluation increased with an increase in the number of generations that make up the phenotypic and genotypic information. In the second trial, the studied methods were compared in an evaluation of progeny without phenotypic. For this purpose, animals of the last three generations (training set) were considered as parents, while with phenotypic and genotypic information, animals of generation 4 (validation set) were considered as progeny. These progenies were found in the genotypic information that was used to determine the allelic relationships, but were not found in the phenotypic information. Therefore, the use of their parents' phenotypic information was evaluated. Using genomic matrices, the accuracy of evaluation increased. Average accuracy of evaluation for each trial was estimated based on 10 iterations, while statistical comparison was performed using student-t test. A significant difference was observed between the evaluation accuracy of the two studied methods.

Key words: Allelic relationships, multiple-trait evaluation, training set, validation set.

INTRODUCTION

It is necessary to create genetic gains in breeding programs, detect animals with superior genetic and select them as parents of the next generation. For this purpose, the breeding value of animals was estimated. During the past decades, genetic gains in quantitative traits made use of genetic selection of parents based on pedigree and phenotypic information (Hill et al., 2008).

Relationship matrix used in traditional BLUP, is formed based on pedigree information. In this method, different relationships information based on their relationship grade with the evaluated animal will take different coefficients. However, using the estimation of the inbreeding and additive genetic relationships, an average value of relationship is considered for all loci, but the Mendelian sampling effect during the formation of gamete, will change the real proportion of genome (Hayes et al., 2006).

The total allelic relationship uses relationship information more precisely, since this method considers the additive genetic variation among families and also uses non-family information of animals. In genetic evaluation programs, information is created through relationship matrix. Therefore, evaluating the use of allelic relationships increases the accuracy of evaluation significantly (Nejati-Javaremi et al., 1997).

In this research, marker genotypes were used with high content of markers to determine allelic relationships,

*Corresponding author. E-mail: Rostami_ankasi@yahoo.com. Tel: +989113951564. Fax: +981912226605.

Table 1. Structure of the simulated base population.

Parameter	Value
Number of animals	100
Number of males	50
Number of females	50
Number of chromosomes	3
Number of markers on each chromosome	200
Number of QTLs on each chromosome	50
Recombination rate	0.005 Morgan
Mutation rate	0.000025

while pedigree relationship matrix was substituted with genomic matrices in the mixed model equations of BLUP. The objective of this research was to study the effect of allelic relationships on the accuracy of multiple-trait genetic evaluation using multiple markers. Then a comparison of the studied methods' accuracy was done in the evaluation progeny without phenotypic records.

MATERIALS AND METHODS

Structure of the population

Firstly, a base population was simulated with the statistical size of 100 members, so that half of these animals were male and others were female. This structure was used in a fixed manner for 110 generations. No selection was done for the convenience of conducting a research, and mating in each generation was performed randomly. In the first trial, to study changes of evaluation accuracy along generations, after creating linkage disequilibrium, the population size of the last generation was expanded 3 times and random mating was done for the next three generations. Then, the phenotypic and genotypic records for females of the last three generations (G1, G2 and G3) were simulated.

In the second trial, the studied methods were compared in evaluation of progeny without phenotypic. For this purpose, animals of the last three generations (training set) were considered as parents with phenotypic and genotypic information, while animals of generation 4 (validation set) were considered as progeny. These progenies were found with the genotypic information that was used to determine allelic relationships, but were not found with the phenotypic information. Therefore, the progeny was evaluated using phenotypic information of their parents.

Structure of the genome

The structure of the base population is shown in Table 1.

Linkage disequilibrium between marker and QTL

The limited size of the effective population (N = 100) created a linkage disequilibrium between the marker and QTL. To calculate r^2, the following formula was used:

$$D = freq(A1\_B1) * freq(A2\_B2) - freq(A1\_B2) * freq(A2\_B1)$$

$$r^2 = D^2 / freq(A1) * freq(A2) * freq(B1) * freq(B2)$$

Genotypes simulation

For this purpose, the effects of QTL for each trait were sampled from the standard deviation of the genotypic value of the loci of that trait (a_sd). The standard deviation of the genotypic value of the second trait (a_sd2) was determined using regression coefficient and this was in the form of a function of the standard deviation of the genotypic value of the first trait (a_sd1), in order to create a genetic correlation among traits. Moreover, the effects of the traits' QTL were sampled using normal distribution.

$a\_sd = 1/10\ \sigma_g$
$cov_{12} = cor * (\sigma_{g1} * \sigma_{g2})$
$b_{2.1} = cov_{12} / v_{g1}$
$a\_sd2 = (b * a\_sd1) / cor$

Phenotypes simulation

Phenotypic records were simulated for the animals of the three last generations. Firstly, the true breeding values (TBV) of the animals were simulated for each trait and were determined in the form of the sum of allelic, locating effects of different loci on each animal. These breeding values were used to determine the phenotypes. For each animal, two phenotypic records (P_1, P_2) were simulated.

$P = TBV + e + \mu$
$e = T z$

$$\mathrm{var}(e) = R \qquad \underset{\sim}{e} \sim N(0, \sigma_e^2)$$

Where, P = phenotypic vector including the records of number of egg and weight of egg for each animal, measuring 2 x 1, TBV = vector of true breeding values of the traits of each animal, measuring 2 x 1, T = Cholesky decomposition of the matrix of environmental (co)variance of the traits, measuring 2 x 2 and Z = vector of random numbers including normal numbers, measuring 2 x 1.

Allelic relationship matrix

To determine the allelic relationships between two animals in a locus, each allele of the first animal was compared with the alleles of the second animal. Subsequently, the extent of similarity of alleles in a locus was determined using the following formula (Nejati et al., 1997):

$$TA = 2 \times \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} I_{ij}}{4} = \frac{\sum_{i=1}^{2}\sum_{j=1}^{2} I_{ij}}{2}$$

Where, I_{ij} shows the similarity of i allelic of an animal with j allele of the second animal. If the two alleles are similar and have the same (I_{ij}), the result will be one, and if they are not the same, the result will be zero. The total allelic similarity for two animals with locus L is

calculated using the following formula (Nejati et al., 1997):

$$TAR = \frac{\sum_{l=1}^{L} TA_l}{L} = \frac{\sum_{l=1}^{L} (\frac{\sum_{i=1}^{2}\sum_{j=1}^{2} I_{ij}}{2})}{L}$$

One of the characteristics of total allelic relationship (TAR) is that a separate allelic relationship matrix is formed for each trait. Contrary to the pedigree method, which uses only one matrix for multiple-trait evaluation, TAR needs three matrices for two-trait genetic evaluation. The matrix of A_{11}^{-1} is used to show the allelic relationships of animals for the first trait, while the matrix of A_{22}^{-1} is used to show the allelic relationships of animals for the second trait. However, the matrix of A_{12}^{-1} is used based on the allele relationships for two traits. To form these allelic relationship matrices, marker information is used.

Firstly, marker effects for each trait were estimated. To estimate such marker effects, the estimation of BLUP was used. Meuwissen et al. (2001) proposed a simple mixed model for the estimation of marker effects and supposed that all markers explain a similar amount of the original variance, due to the fact that the variance of each locus is σ^2_G/n. Here, σ^2_G is the total genetic variance and n is the total number of markers. The statistic model used to estimate marker effects is as follows:

$$y = X b + Z m + e$$

In this model, y = vector of the observations, b = vector of fixed effects (mean) and m = vector of marker effects. X and Z are the matrices of the coefficients relating to the records of mean and marker effects, respectively. The elements of Z matrix are 0, 1 and 2, which are the signs of homozygosis for the first allele, and heterozygosis and homozygosis for the second allele, respectively.

In this study, the mean and variance of observations were considered as 1 μ and $V(y) = ZIZ'\sigma_m^2 + I\sigma_e^2$, respectively. Henderson mixed equations were formed to estimate marker effects.

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

The single-trait evaluation was conducted once for the number of egg and in another time for the weight of egg in order to determine the marker effects for both traits. The marker, whose effect for both traits was more than 2% of the genetic standard deviation of the considered traits, was considered a common marker. If the effect of the marker was more than 2% of the genetic standard deviation for only one trait, this marker was considered the specific marker of that trait. Otherwise, it is considered as a neutral marker. So, the specific and common markers of traits were determined, and the matrices of allelic relationship were formed. In forming the matrix of A_{11}^{-1}, the markers of trait 1 and the common markers of both traits were used, while in forming the matrix of A_{22}^{-1}, the markers of trait 2 and the common markers of both traits were applied and in Forming the matrix of A_{12}^{-1}, the common markers of both traits were used.

Table 2. Parameters used for the considered traits.

Parameter	Trait			
	μ_p	Var_g	Var_e	h^2
Number of egg (tr_1)	38	28.08	149.08	0.159
Weight of egg (tr_2)	47	7.3	8.32	0.467

$$LHS = \begin{bmatrix} X'_1R^{11}X_1 & X'_1R^{12}X_2 & X'_1R^{11}Z_1 & X'_1R^{12}Z_2 \\ X'_1R^{11}X_1 & X'_2R^{22}X_2 & X'_2R^{21}Z_1 & X'_2R^{22}Z_2 \\ Z'_1R^{11}X_1 & X'_1R^{12}X_2 & Z'_1R^{11}Z_1+A_{12}^{-1}g^{11} & Z'_1R^{12}Z_2+A_{12}^{-1}g^{12} \\ Z'_2R^{21}X_1 & Z'_2R^{22}X_2 & Z'_2R^{21}Z_1+A_{12}^{-1}g^{21} & Z'_2R^{22}Z_2+A_{22}^{-1}g^{22} \end{bmatrix}$$

RESULTS

In this study, multiple-trait evaluation of the number and weight of egg was done for traits. For this purpose, the published information for Iranian native fowls was used. The parameters used for these traits were shown in Table 2 (Ghazikhani et al., 2007).

Genetic and environmental covariance between the considered traits was -3.8 and 0.708, respectively and their genetic correlation was equal to -0.2654139. Upon formation of G and R matrices, their covariance was formed in order for it to be used in the evaluation with BLUP mixed equations. The matrices of G^{-1} and R^{-1} are as follows:

$$G^{-1} = \begin{bmatrix} 0.03831136 & 0.0199429 \\ 0.0199429 & 0.14736754 \end{bmatrix}$$

$$R^{-1} = \begin{bmatrix} 0.00671052 & -0.00057104 \\ -0.00057104 & 0.1202409 \end{bmatrix}$$

The number of simulated specific and common genes is shown in Table 3.

The changes of evaluation accuracy of the considered traits across generations using the methods of TAR–BLUP and TP–BLUP are shown in Table 4. Consequently, the phenotypic and genotypic records for females of these three generations (G1, G2 and G3) were simulated. Breeding values were estimated using information from the current and all previous generations, but correlations of estimated breeding values (EBV) with true breeding values were calculated within generations.

The results showed that the accuracy of evaluation increased with the increase in the number of generations who made the information. Correlations of BV with EBV are a function of heritability and the amount of information available, increases across generations. Also, the accuracy of evaluation increased with the increase of

Table 3. Number of simulated specific and common genes.

Parameter	Value
Number of specific genes of trait 1	21
Number of specific genes of trait 2	18
Number of common genes	106

Table 4. Values of evaluation accuracy of the considered traits across the three generations.

Generation	TP-BLUP		TAR-BLUP	
	Egg number	Egg weight	Egg number	Egg weight
1	0.40863±0.015	0.64129±0.011	0.50005±0.017	0.72984±0.018
2	0.42998±0.017	0.65999±0.019	0.53129±0.019	0.73885±0.011
3	0.43880±0.020	0.67681±0.018	0.55807±0.010	0.75990±0.023

Table 5. Values of evaluation accuracy of progeny without phenotypic information for the considered traits.

Generation	TP-BLUP		TAR-BLUP	
	Egg number	Egg weight	Egg number	Egg weight
Parents	0.4258±0.015	0.65936±0.017	0.5298±0.019	0.74286±0.015
Progeny	0.1875±0.012	0.39625±0.013	0.35102±0.018	0.59338±0.015

heritability (Villumsen et al., 2009; Goddard, 2009).

Then the animals of the last three generations (G1, G2 and G3) were considered as parents and were evaluated based on all the present information. Animals of G4 were without phenotypic records, but their genotypic information was available. These progenies were evaluated using the phenotypic information of their parents.

Accuracy of evaluation of progeny without phenotypic information by methods of TAR – BLUP and TP – BLUP has is shown in Table 5. The TAR-BLUP method used genotypic information of progeny to form genomic matrices. Therefore, accuracy of evaluation increased using TAR-BLUP.

Meuwissen et al. (2001) simulated a training set with 500 phenotypic records and estimated the genomic breeding values of the animals of validation set. The accuracy of the single-trait evaluation for trait with high heritability was 0.579.

Solberg et al. (2008) used 1000 phenotypic records from training set and evaluated the animals of validation set using Bayes-B method. When the Bayes method was considered, the accuracy evaluation was 0.663.

In this study, accuracy of evaluation of animals without phenotypic information was estimated using multiple-trait TAR-BLUP. Accuracy of the evaluation for traits with high and low heritability was 0.59 and 0.35, respectively.

The accuracy of evaluation in TP-BULP is the function of heritability and the available phenotypic information which increases across generations. When the genotypic information was considered, the accuracy of evaluation with TAR-BLUP was always bigger than TP-BLUP.

In this study, the marker effects were estimated using a simple mixed model of BLUP and were determined for each trait. Then, the markers were grouped according to these effects, while particular markers of each trait, as well as common markers among two traits, were determined and used to form the allelic relationship matrices. These matrices were converted and used in multiple-trait genetic evaluation using the mixed model of BLUP. Nevertheless, the Bayes methods estimate the marker effects more accurately (Meuwissen et al., 2001; Solberg et al., 2008; Hayes et al., 2009).

Total allelic relationship uses, more precisely, information from records of relatives because it accurately accounts for the variation in the additive genetic relationship among relatives and also because it uses information from pedigree-unrelated individuals, which leads to increased accuracy. Using traditional BLUP in multiple-trait evaluation, a similar relationship matrix is used for all traits, whereas animals have different allele relationships for various traits. Consequently, in total allelic relationships, separate matrices, which are an advantage of applying genomic matrices, are formed.

Therefore, multiple-trait evaluation through TAR-BLUP is suggested for traits with low heritability, traits which are limited to sex and traits whose measurement is difficult. Considering the genotypic information, the accuracy of evaluation was improved.

REFERENCES

Ghazikhani Shad A, Nejati Javaremi A, Mehrabani Yeganeh H (2007). Animal model estimation of genetic parameters for most important economic traits in Iranian native fowls. Pak. J. Biol. Sci. 10(16): 2787-2789.

Goddard ME (2009). Prediction of accuracy and maximization of long term response. J. Genet. 136: 245-257.

Hayes BJ, Chamberlain A, Goddard ME (2006). Use of linkage markers in linkage disequilibrium with QTL in breeding programs. Proc. 8th World. Congress of Genet. Belo Horizonte, Brazil.

Hayes BJ, Bowman PJ, Chamberlain A, Goddard ME (2009). Genomic selection in dairy cattle. J. Dairy Sci. 92: 433-443.

Hill WG, Goddard ME, Visscher PM (2008). Data and theory point to mainly additive genetic variance for complex traits. J. Genet. 4: 100-108.

Meuwissen THE, Hayes BJ, Goddard ME (2001). Prediction of total genetics value using genome-wide dense marker map. J. Genet.177: 1819-1829.

Nejati Javaremi A, Smith C, Gibson JP (1997). Effect of total allelic relationship on accuracy and response to selection. J. Anim. Sci. 75: 1738-1745.

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008). Genomic selection using different marker types and densities. J. Anim. Sci. 86(10): p. 2447.

Villumsen TM, Janss L, Lund MS (2009). The importance of haplotype length and heritability using genomic selection in dairy cattle. J. Anim. Breed. Genet. 126: 3-13.