

GAIN IN ACCURACY OF EVALUATION BY INCLUDING TOTAL ALLELIC IDENTITY

A. Nejati-Javaremi, J.P. Gibson and C. Smith

Center for Genetic Improvement of Livestock, Department of Animal and Poultry Science, University of Guelph, Guelph, Ontario, N1G 2W1
Canada.

SUMMARY

Alleles at a locus in an individual can be identical by descent or be alike in state. Adding both of these for all segregating loci gives the total allelic identity for a trait. It is shown by stochastic simulation of unselected population that estimated breeding values derived using a relationship based on total allelic identity (EBVT) are more closely correlated with true breeding value (TBV) than conventional estimated breeding values (EBVR) derived using the estimated additive relationship based on pedigree. Correlations of TBV with EBVT were 0.903, 0.865 and 0.664 respectively for 5, 10 and 100 loci simulated, compared with 0.552, 0.598 and 0.609 with EBVR. With the advent of large numbers of genetic markers, and the identification of trait loci, use of total allelic identity should be considered.

INTRODUCTION

The degree of genetic identity among animals has traditionally been measured using pedigree information. This measure of identity is an average value and has variation associated with it. For example the pedigree derived additive genetic relationship of a parent with its progeny is considered to be the same as that of two full sibs. In the absence of inbreeding and when parents are unrelated, the former true relationship is 0.5 but the latter relationship can vary from zero to one. The variance of true relationships about the pedigree prediction has been investigated for a few types of relationships (Hill 1993a, Hill 1993b, Rasmuson 1993). The impact of such variation on the use of relationship information has not been investigated.

The variance in identity is dependent on the number of segregating loci. An infinitesimal model of gene action usually is considered in predicting relationship, whereas it is likely that a finite number of loci are responsible for a considerable portion of the variation in a trait.

In predicting the relationship among animals it is usually assumed that the individuals of the base population have zero inbreeding and are unrelated. By

substituting total allelic identity the structure of the matrix of additive genetic relationship in the base generation and after will change.

The traditional matrix of additive genetic relationships is commonly used in mixed model equations to incorporate information from relatives. Molecular genetics is adding to our knowledge of genetic constitution of animals and plants. It may become economical to measure the proportion of identical genes among individuals. Nongenetic residual effects make it difficult to assign the correct additive value to individual genes when the number of loci is not small but finite. Using the information of total allelic identity in genetic evaluation may increase accuracy of selection. This study is the first of a series on the impact of using total allelic identity.

MATERIALS AND METHODS

A population of size 20 (10 of each sex) was simulated. Genetic variation was caused by a finite number of loci, each with 2 alleles, with the additive genetic effect of loci sampled from a normal distribution, with mean $\mu = 0$ and variance σ_{ai}^2 , and effects scaled so that the total additive genetic variance was σ_a^2 ($\sigma_a^2 = L\sigma_{ai}^2$, where L = number of loci). The heritability was set to 0.25. Random mating was practiced for 5 generations. Each individual was mated to two individuals of the opposite sex with each mating giving one offspring. Random residual effects were added to true additive genetic effects of individuals to produce a phenotypic record for each individual. There were 5, 10 or 100 loci. No selection was applied.

Best Linear Unbiased Prediction (BLUP)(Henderson 1972) using an animal model and standard mixed model equations was used to estimate breeding values (EBV) of individuals based on the phenotypic records and different relationship matrices. The linear model was:

$$y_{ij} = g_i + u_{ij} + e_{ij}$$

Where:

- y_{ij} is the phenotypic record.
- g_i is the mean of the generation i .
- u_{ij} is the additive genetic effect of animal j of generation i .
- e_{ij} is a random residual.

Individual records were assumed to be sampled from a large population with mean equal to zero. Due to finite numbers in each sample, however, the averages of the base and of subsequent generations fluctuate randomly around the population

mean. Therefore the fixed effect of generation mean is fitted in the model to account for the deviation of each generation's mean from zero.

It is further assumed that the actual heritability of each sampled population is equal to the base heritability of 0.25. This is not true for individual populations but will hold on average over multiple sampling (100 replicates of simulation). The variance of heritability about the expectation increases with a decreasing number of loci.

Two approaches were taken to compute the relationship among individuals:

1) The traditional method of using pedigree information to predict identity by descent was used to predict the proportion of genes each pair of individuals have in common due to having common ancestors (Wright 1922).

2) Every pair of individuals were compared at the genotypic level and the proportion of shared alleles derived. Total allelic identity (TA) between two individuals at locus l was calculated as,

$$TA_l = \frac{\sum_{i=1}^2 \sum_{j=1}^2 I_{ij}}{2},$$

where I_{ij} is the identity of the i^{th} allele of the first individual with the j^{th} allele of the second, and takes the value of 1 if the two alleles are identical and zero if they are not. Total allelic identity of individuals i and j is then:

$$TA_{ij} = \frac{\sum_{l=1}^L TA_l}{L},$$

where L is the total number of loci.

RESULTS

Table 1 gives correlations of phenotypes, estimated breeding values using identity by descent (EBVR) and estimated breeding values using total allelic identity (EBVT) with true breeding values (TBV). Each correlation is the average of 100 replicates of simulation. Average observed heritability, derived using the variances of TBV and of residual effects, is also presented.

Table 1. Correlations of phenotypes, EBVR and EBVT with TBV (Average of of 100 replicates across 5 generations). Values in the brackets are standard errors.

Number of Loci	Correlation with TBV			Average observed h^2
	Phenotype	EBVR	EBVT	
5	0.435(0.0137)	0.552(0.0153)	0.903(0.0089)	0.21(0.011)
10	0.491(0.0112)	0.598(0.0138)	0.865(0.0102)	0.23(0.009)
100	0.489(0.0085)	0.609(0.0098)	0.664(0.0101)	0.25(0.005)

DISCUSSION

Clearly the correlations of EBVT with TBV are higher than those for EBVR. The difference decreases with increasing number of loci. The pedigree estimate of identity by descent is an appropriate method of accounting for relationships when the infinitesimal model holds. But it is clear that using total allelic identity can increase the accuracy of selection. Even with 100 loci there was nearly a 10% gain in accuracy. These results are from a population structure where most of the information comes from the individual, its parents and progeny and two half-sibs, where the variance of relationship is small compared to that with a full-sib family structure (unpublished results). There is also evidence that most of the information used by EBVT comes from the genotype of the individual and its relatives but phenotype is a major contribution to EBVR. Part of the gain in accuracy is due to accounting for the total allelic identity among individuals of the base generation but the major part is due to accounting for variation in total allelic identity across generations.

Obviously, true total allelic identity can not be known in practice. So the present results set an upper limit to the gain in accuracy possible. Information on a large numbers of molecular marks (Rohrer *at al.* 1994) can estimate the total allelic identity for a trait. Research is currently under way to determine how much information can be gained by use of marker information.

REFERENCES

- HENDERSON, C.R. (1972) In Proceedings of the Animal Breeding and Genetics Symposium in honor of Dr. Jay L. Lush.
- HILL, W.G. (1993a) *The Journal of Heredity* 84: 212-213.
- HILL, W.G. (1993b) *Heredity* 71: 652-653.
- RASMUSON, M. (1993) *Heredity* 70: 266-268.
- ROHRER, G.A., ALEXANDER, L.J., KEELE, J.W., SMITH, T.P. and BEATTIE, C.W. (1994) *Genetics* 136: 231-245.
- WRIGHT, S. (1922) *American Naturalist* 56: 330-338.