ESTIMATING GENETIC MERIT WHEN GENOTYPE DATA ARE INCOMPLETE

Bruce Tier and Kim Bunter
Animal Genetics and Breeding Unit[1], University of New England, Armidale NSW Australia 2351.

SUMMARY
Direct markers for QTL are becoming available and, for traits that are difficult to measure or observed later in life, offer the potential to increase the accuracy of EBVs. While genotyping costs are non-trivial, QTL information will not be available for all individuals – with important consequences for genetic evaluation systems. Using simulated populations in which all parents and some progeny are genotyped for a QTL acting additively, two methods of using genotypic information are compared with EBVs based on the infinitessimal model. Both methods – deregressing EBVs for the QTL effect according to their accuracy, or using heterogeneous residual variances – improve correlations between simulated and estimated genetic merit, particularly for individuals without performance data.
Keywords: QTL, direct marker, genetic evaluation

INTRODUCTION

Many traits of economic importance are controlled by a large number of genes (polygenes) acting in concert. Selection on estimated breeding values (EBVs) based on the infinitessimal model using Best Linear Unbiased Prediction (BLUP) has proven to be very effective for traits that are easily measured. Recently, much research effort has been applied to finding genes with a large effect– so-called quantitative trait loci (QTL) – on quantitative traits. QTL and/or markers linked to QTL have been discovered for most livestock species. Given the effectiveness of selection based on current methods (e.g. BLUP) there is a general consensus that QTL are only going to be of use when the traits are expensive to measure, they are expressed later in life or they are sex-limited. In these cases significant improvement in the accuracy of estimated genetic merit are to be expected from genotyping individuals for the QTL.

Tests for QTL are becoming available for many livestock species (e.g. marbling in beef cattle). However, typically only a relatively small number of animals, as a proportion of the population, are genotyped. This will continue until the cost of genotyping reduces. In the meantime genotyping will likely be limited to animals of importance, that is mainly current and/or prospective parents.

Methods for performing marker assisted selection (MAS) have been suggested by various authors (e.g. Fernando and Grossman 1989). These methods generally assume that genotypic data are available for all individuals. Data augmentation can be used to impute the missing genotypic data in a Bayesian context (Hoeschele 2001), but this would require MCMC methods which, for routine evaluation of large populations, are currently impractical. Alternatively, methods exist for inferring genotype probabilities for ungenotyped individuals given the pedigree of the population and the known genotypes. These methods only provide certainty for progeny of homozygous parents and for

[1] AGBU is a joint unit of NSW Agriculture and the University of New England

parents of progeny carrying both alleles. Also, variances differ between genotyped and ungenotyped individuals when the QTL has a significant effect. The effect of moderately sized QTL acting additively has been found to be incorporated into estimates of polygenic effects (Tier and Henshall, 2000).

This paper compares alternative methods of using partial genotypic information in routine genetic evaluation. These include 1) modifying EBVs after the population has been evaluated without any genotypic information, and 2) assuming heterogeneous residual variances for different classes of individuals, according to their genotypic information.

MATERIALS AND METHODS

A series of populations were simulated, modeled on a sheep population with 200 ewes mated to 8 sires in each of 10 years. Replacement parents were randomly chosen from among the progeny. 50% of sires and 25% of dams were replaced each year. Different levels of QTL effects (5 and 10% of the total variance) and different polygenic proportions (10% and 33% of the total variance) were simulated. Data (y) were generated using a model: y=b+a+q+e, where b is the mean of the contemporary group, a is the polygenic effect, q is the effect of the QTL and e is a residual. Records were generated for all non-founder individuals – founders were unobserved. For the lower heritability option data were also limited to female parents. The QTL acted additively, was inherited according to mendelian sampling and founder alleles had an equal chance of being A or B. There were three genotype classes AA, AB and BB with effects of y, 0 and $-y$ respectively. All parents were genotyped. Where possible – for progeny of homozygous parents – genotypes were inferred. Different proportions (10, 30, 100%) of the remaining, uninferrable, progeny were randomly chosen to be genotyped. The genotype probabilities of progeny from other crosses are shown in Table 1. 100 replicates for each combination of effects were simulated.

Table 1: Genotype probabilities, and within family means and variances resulting from all possible crosses in a two allele locus.

Parental genotypes		Within family genotype probabilities			Within family statistics	
Parent 1	Parent 2	AA	AB	BB	Mean	Variance
AA	AA	1.0	-	-	y	0
AA	AB	0.5	0.5	-	$y/2$	$y^2/4$
AA	BB	-	1.0	-	0	0
AB	AB	0.25	0.5	0.25	0	$y^2/2$
AB	BB	-	0.5	0.5	$-y/2$	$y^2/4$
BB	BB	-	-	1.0	$-y$	0

These data were analysed with three different methods. The first method (infinitessimal) used a typical infinitessimal model to evaluate the animals. Both genetic effects – the polygenic and the QTL – were included in a single breeding value (u=a+q). The variance of the true breeding value was the sum of the variances due to the polygenic and QTL effects (Var(u)=Var(a)+Var(q)). The second method (deregressed) adjusted the EBVs obtained from the first method using the formula

EBV*=EBV+(1-acc^2)q*, where acc is the accuracy of the EBV derived from the first model and q* is the effect of the QTL determined by the animal's genotype if known or its parents' genotypes – the within family mean in Table 1 – if not. The third method (heterogeneous) fitted the polygenic and QTL effects independently. Different mean effects and residual variances were used depending upon each individual's genotypic status. Data were pre-adjusted for the QTL effect if known, otherwise according to the family means shown in Table 1. The variance of the polygenic effects was the simulated value (Var(a)). The residual variance was augmented by the appropriate within family variance when the QTL genotype was unknown (Table 1). The value of the QTL effect was added to the polygenic EBV to give an estimate of each animal's genetic merit. Estimates of genetic merit from the three evaluation methods were compared with simulated values.

Table 2: Correlations (empirical standard errors) between estimated and simulated genetic merit for different models, levels of heritabilities, additive QTL effects, potential phenotypes and proportions of progeny genotyped in the sample populations– means of 100 replicates.

Data descriptors						
Var(q)	0.05			0.1		
Var(a)	0.33	0.1	0.1	0.33	0.1	0.1
Phenotypes	All progeny	All progeny	Female parents	All progeny	All progeny	Female parents
Evaluation method						
Infinitessimal	0.716 (0.030)	0.573 (0.050)	0.303 (0.082)	0.739 (0.027)	0.615 (0.043)	0.264 (0.082)
10% of ambiguous progeny genotyped						
Deregressed	0.727 (0.029)	0.648 (0.040)	0.546 (0.050)	0.757 (0.025)	0.707 (0.031)	0.601 (0.042)
Heterogeneous	0.731 (0.028)	0.659 (0.036)	0.575 (0.044)	0.761 (0.024)	0.721 (0.028)	0.656 (0.034)
30% of ambiguous progeny genotyped						
Deregressed	0.731 (0.028)	0.663 (0.038)	0.586 (0.044)	0.762 (0.025)	0.726 (0.029)	0.660 (0.034)
Heterogeneous	0.731 (0.028)	0.674 (0.035)	0.599 (0.042)	0.760 (0.024)	0.742 (0.026)	0.688 (0.030)
All progeny genotyped						
Deregressed	0.739 (0.027)	0.697 (0.033)	0.638 (0.037)	0.774 (0.023)	0.768 (0.024)	0.729 (0.027)
Heterogeneous	0.730 (0.028)	0.708 (0.031)	0.651 (0.034)	0.764 (0.024)	0.789 (0.022)	0.756 (0.023)

RESULTS AND DISCUSSION

At the higher level of polygenic variance there is little to choose between either of the methods that use QTL information, both of which are only slightly better (≤5% increase in accuracy) than using the infinitessimal model when pheontypes are recorded. When the whole population is considered

little is gained from genotyping more than 10% of the population although, with all parents genotyped and the gene frequency at 0.5, the genotypes of approximately 50% of all progeny can be inferred. However, when only the most recent cohort is considered, the benefit of genotyping more progeny (not shown) approaches significance.

At the lower polygenic variance, when all progeny have phenotypes, estimates of genetic merit derived from the deregressed method are slightly, but not significantly, less correlated with true merit than those derived from the model which fits the genetic effects separately. Both methods provide more accurate and less variable estimates of the animals' genetic merit than the infinitessimal model. The benefit of using either deregressed or heterogenous methods compared with the infinitessimal method is much more pronounced when data are only available on the female parents. The benefit of using methods that exploit the QTL increases with the size of its effect.

When data are only available on dams, the correlation between simulated and predicted genetic merit for the infinitessimal model is 0.303 when Var(q) is 0.05 and 0.264 when Var(q) is 0.1. This is the only instance when the QTL with a smaller effect induces a higher correlation between estimated and simulated merit than the QTL with the larger effect, under otherwise similar conditions. While this difference is small it is usually in the other direction. This, and the greater variation of the results, suggests that the infinitessimal model is inefficient when the QTL is generating a large proportion of the total genetic variation and genotypic information should be used when evaluating genetic merit.

It is unlikely that a different sized population or mating structure will produce radically different results. The effect of selection is likely to lead to an increase in the proportion of homozygous parents, and a consequent increase in the quantity of progeny whose genotypes can be inferred. The effect of selection and alternative modes of gene action on the predictions of genetic merit by deregressing EBVs generated ignoring any genotypic information are yet to be tested. Similarly, alternative strategies for analysing populations for multiple traits and with different genotyping strategies – such as all or current sires only – need consideration.

CONCLUSION

Deregressed EBVs from an infinitessimal model appear to be as useful predictors of genetic merit as those derived from a model using heterogeneous residual variances to accommodate differences in the available genotypic information. The former is simpler to implement and suggests a first approach to using limited genotype data in genetic evaluation systems. Ignoring the genotypic data limits the accuracy of selection, particularly as the effect of the QTL increases.

REFERENCES

Fernando R.L., Grossman M. (1989) *Genet. Sel.Evol.* **21**:467-477.

Hoeschele I. (2002) In "Handbook of Statistical Genetics", p. 599. editors Balding D.J., Bishop M and C. Cannings (2001) Wiley. New York.

Tier B., Henshall J.M. (2001) *Genet. Sel. Evol.* **33**:587-603.