INFERENCE ON GENOTYPE PROBABILITY BASED ON POLYGENIC ESTIMATED BREEDING VALUES

A. Reverter and B. Tier

Animal Genetics and Breeding Unit , University of New England and NSW Agriculture, Armidale, NSW 2351, Australia

INTRODUCTION

The majority of the existing programs for national genetic evaluation of livestock species compute estimated breeding values (EBV) assuming a standard infinitesimal model ignoring the presence of any major genes (QTL) because such models have proven to be useful and robust in practice. More realistic models do not necessarily provide better predictions if their parameters cannot be estimated accurately (Goddard, 2001). Therefore, the methodology used to compute EBV is expected to remain in used in the near future. However, with an increasing use of QTL by animal breeders, there is the need to integrate traditional methodology with new genetic technology. Once a QTL has been identified, there is a potential for substantial savings if only a proportion of animals are genotyped while maximizing the utility of the resulting information in a marker assisted selection scheme (Kinghorn, 1999).

An EBV computed ignoring the existence of a QTL is likely to capture most of the QTL effect particularly if the effect is completely additive. The objective of this study is to evaluate the conditional probability of the QTL genotype for an individual given its EBV. Expectations are developed and their value investigated using stochastic simulation.

MATERIAL AND METHODS

Expectations. The trait under consideration is assumed to be genetically controlled by an infinite number of additive loci, each with an infinitesimal polygenic effect plus a single biallelic locus (alleles A and B) of large effect. The conditional probability of the QTL genotype (either AA, AB, or BB) given the EBV is given by:

$$P[QTL / EBV] = \frac{P[QTL]\,P[EBV / QTL]}{P[EBV]}$$

where P[EBV] is assumed normal with zero mean and variance σ^2_{EBV}; P[QTL] is assumed trinomial with probability p^2, $2pq$, and q^2 for QTL genotype AA, AB, and BB, respectively, and where p is the frequency of the favourable allele (A) and $q = (1 - p)$ the frequency of allele B. The variance explained by the QTL is $\sigma^2_{QTL} = 2pq\alpha^2$, and α is the average effect of the gene substitution. The conditional probability of an EBV given the QTL (P[EBV/QTL]) is unknown. Hayes and Goddard (2001) suggested a gamma density of QTL effects which is likely to be approximated by a normal distribution provided a single QTL operates and the number of individuals within each QTL genotype is sufficiently large. An approximation to the mean and variance is given by:

$$E[EBV/QTL] = rh\left(\alpha + \frac{\alpha}{\sigma_{EBV}}\right) = \frac{\alpha}{\sigma_P}(r\sigma_G + 1) = \frac{\alpha}{\sigma_P}(\sigma_{EBV} + 1),$$

$$Var[EBV/QTL] = \sigma^2_{EBV} - Var\left[\frac{\alpha}{\sigma_P}(\sigma_{EBV} + 1)\right] = \sigma^2_{EBV} - h^2_{QTL}(\sigma_{EBV} + 1)^2 ,$$

where σ_P is the phenotypic SD, h is the square root of the total heritability (polygenic plus QTL) used to compute the EBV, h^2_{QTL} is the QTL heritability, and r is the accuracy of EBV prediction given by σ_{EBV}/σ_G where σ_G is the additive genetic SD. The above expressions result from the assumption that the QTL effect at the phenotypic level equals that at the genetic level and from a well-known result of conditional first and second moments such that Var(x) = E[Var(x/y)] + Var[E(x/y)] for two random variables x and y. With perfect accuracy (i.e., r = 1.0 and $\sigma_{EBV} = \sigma_G$), the QTL effect in σ_P units equal that in σ_G units and the variance of the EBV given the QTL equals σ^2_G adjusted for the h^2_{QTL}. At a given QTL variance, a lower frequency for the favourable allele implies a QTL with large effect. In situations where the QTL accounts for the majority of σ^2_G, the variance of the EBV within each QTL genotype will be vastly reduced.

Simulation. Data were simulated from a foundation of 40 males and 1000 females. Mating was random with five overlapping generations until 6040 individuals were produced. Replacement rates were 80% for males and 20% for females and one offspring was generated per mating. Phenotypes were simulated with $\sigma^2_P = 100$ with total h^2 (polygenic plus QTL) of 30%. Polygenic values for foundations were obtained from a normal distribution with zero mean and variance σ^2_G. The alleles at the QTL were chosen at random with probability p of allele A. In subsequent generations, polygenic values for the offspring were generated as the mid-parental polygenic value plus a Mendelian deviation which took account of inbreeding. The QTL genotype was obtained by random sampling of one allele from each parent. The fixed effect of sex with two levels was also simulated. The QTL was simulated without dominance.

Four cases were investigated, corresponding to all combinations of p (01 or 05) and σ^2_{QTL} (3 or 27). These four cases correspond to a QTL effect in σ_P units of 0.408, 0.245, 1.225, and 0.735, respectively for cases 1 to 4. For each case, a total of 100 replicates were simulated.

Validation. For each simulated data set, genetic parameter estimates and EBV were obtained from a standard mixed model equation including sex as a fixed and additive genetic as a random effect using the VCE software of Groeneveld and García-Cortés (1998). The mean and variance of the resulting EBV were evaluated for the whole data and within each genotype (AA, AB, or BB). Empirical values were contrasted with those from the proposed equations.

RESULTS AND DISCUSSION

Table 1 shows the means and variances for the EBV from the whole data set and within each genotype after 100 replicates of simulation along with the expectations obtained from the

proposed equations. In all cases, there were no significant differences between the expectations for the mean EBV given the QTL and those empirically observed (1.97 *vs* 2.03, 1.18 *vs* 1.18, 5.86 *vs* 5.94, and 3.52 *vs* 3.55, for cases 1 to 4, respectively). Similarly, the expectations for the variance of the EBV given the QTL were not different to those observed (14.06 vs 14.00, 13.92 vs 13.88, 8.23 vs 8.15, 8.25 vs 8.16, for cases 1 to 4, respectively). However, these comparisons must be taken with caution as simulation results are influenced by the number of individuals within each genotype. This was particularly apparent with $p = 0.1$ (cases 1 and 3) where the number of individuals with genotype AA ranged from 18 to 121 and the STD of the mean EBV where two and five times as large as that for genotypes AB and BB, respectively.

Table 1. Means and variances for the EBV from the whole data set and within each genotype after 100 replicates of each case[A]

	Case 1	Case 2	Case 3	Case 4
p = P(A)	0.10	0.50	0.10	0.50
Var(QTL)	3	3	27	27
EBV whole data set				
Mean	-0.05(0.28)	0.00(0.27)	-0.03(0.27)	-0.02(0.28)
Variance	14.77(2.20)	14.62(2.33)	14.44(2.51)	14.48(1.80)
EBV given QTL = AA				
Mean	3.60(0.94)	1.18(0.36)	10.67(1.12)	3.55(0.33)
Variance	13.39(4.22)	13.92(2.07)	8.07(3.46)	7.87(0.94)
EBV given QTL = AB				
Mean	1.57(0.40)	0.01(0.27)	4.69(0.46)	-0.01(0.22)
Variance	13.93(2.36)	13.94(2.14)	9.30(2.08)	8.39(0.91)
EBV given QTL = BB				
Mean	-0.45(0.29)	-1.18(0.45)	-1.21(0.24)	-3.56(0.37)
Variance	14.02(2.15)	13.70(2.41)	7.90(0.98)	8.00(0.96)
Expectations				
E[EBV/QTL]	1.97(0.10)	1.18(0.07)	5.86(0.36)	3.52(0.18)
Var[EBV/QTL]	14.06(2.11)	13.92(2.24)	8.23(1.66)	8.25(1.18)

[A]In all cases, phenotypic variance was 100 and total heritability (polygenic plus QTL) was 30%. Values in brackets correspond to the SD after 100 replicates.

Figure 1 illustrates the genotype probabilities as a function of the magnitude of the EBV in standard deviation units and for each case. An individual with and EBV greater than 2 σ_{EBV} units will have a QTL genotype AA with probability greater than 4%, 39%, 7% and 79% for cases 1 to 4, respectively.

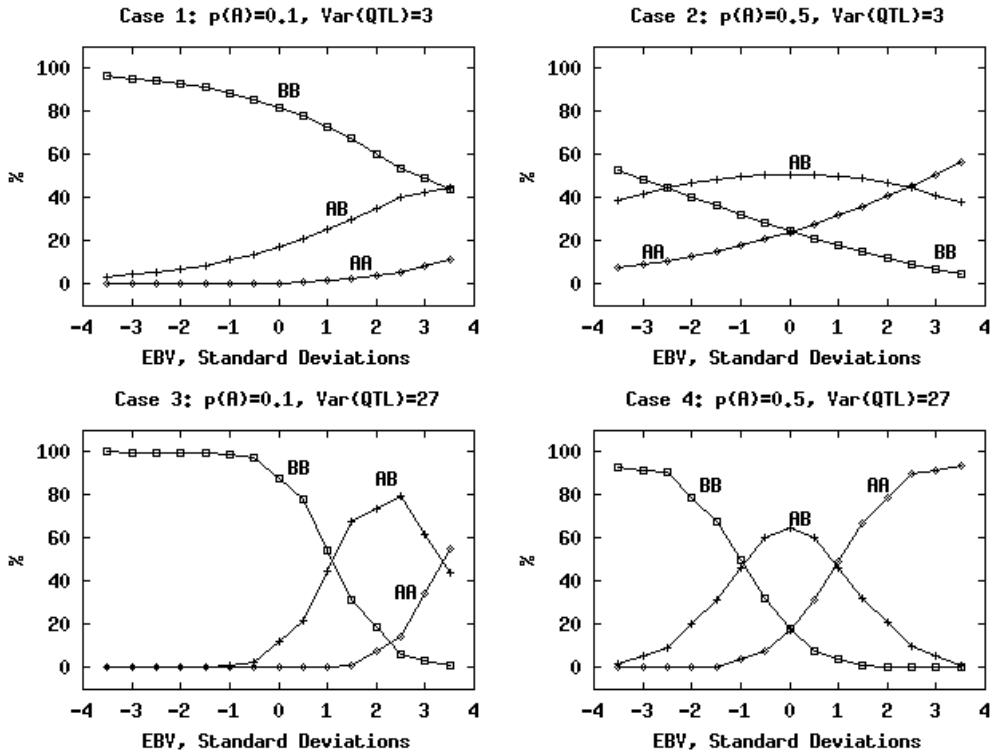

Figure 1. Genotype probabilities as a function of the magnitude of the EBV in standard deviation units and for each case

CONCLUSION

Unlike existing methods for determining genotype probabilities (see Tier and Henshall (2001) for a recent algorithm), the proposed method is not limited to known pedigree structures or reliant on a seemingly large proportion of animals having been genotyped to provide useful results. However, this method is not meant to compete with existing ones but rather to complement them in situations when EBV are available. Further research is required to test this procedure with dominance and selection.

REFERENCES

Goddard, M.E. (2001) *Livest. Prod. Sci.* **72** : 117-127.
Groeneveld and García-Cortés (1998) *Proc. 6th WCGALP* **27** : 455-456.
Hayes, B.J. and Goddard, M.E. (2001) *Genet. Sel. Evol.* **33** : 209-229.
Kinghorn, B.P. (1999) *J. Anim. Breed. Genet.* **116** : 175-180.
Tier, B. and Henshall, J. (2001) *Genet. Sel. Evol*. **33** : 587-604.