PREDICTION OF SINGLE LOCUS EFFECTS USING GAMETIC MODELS

M. H. Yazdi and L. R. Schaeffer

CGIL, Animal & Poultry Science, University of Guelph,
Guelph, ON, Canada N1G 2W1

INTRODUCTION

Best linear unbiased prediction (BLUP) methods are currently applied to the prediction of breeding values of animals based on phenotypes of the animals and their relatives. The infinitesimal animal model assumes a large number of independent additively genetic loci affecting the phenotypes of the animals and requires that the pedigree information be complete for the unbiased estimation of breeding values. Equivalently, a gametic model (two gametes per animal) can be used to predict breeding values as well, although the number of equations for solving is almost twice as many as in an animal model. Schaeffer et al. (1989) and Jamrozik and Schaeffer (1991) show how the gametic relationship matrix and its inverse are constructed, and the equivalence of gametic models to an animal model. Gametic models have been applied to the analysis of gametic imprinting effects (Gibson et al., 1988; Tier and Solkner, 1993). This paper demonstrates how gametic models can be used to estimate single locus effects. Much use is made of the results from van Arendonk et al. (1994), but in a more simplified procedure.

METHODS

Single Locus Relationship Matrices. This section extends Henderson (1985) to the prediction of single locus effects. Suppose that a phenotypic trait is controlled by one locus with several alleles and that all animals are genotyped for that locus. Van Arendonk et al. (1994) showed how to construct a gametic relationship matrix for a single locus thereby tracing the alleles of each parent to all progeny (p 321 with r=0.0). Let this matrix be denoted as $H_i$ for the $i^{th}$ locus. This matrix is generally singular (i.e. no inverse). Associated with this locus is a genetic variance, $\sigma_i^2$.

In an infinitesimal model, all loci are assumed to be independent in a non-selected population. Conceptually, a matrix, $H_i$, could be constructed for each locus. Applying the independence assumption, then the sum from one to infinity of all $H_i\sigma_i^2$ matrices should be equal to the total gametic relationship matrix ($G(0.5)\sigma_a^2$) described by Smith and Allaire (1985). Combining this observation with Henderson (1985), then the total gametic effects, ($g$ with two per animal), could be predicted from a gametic model. Following this, the genetic effects for the $i^{th}$ locus, $h_i$, could be estimated as $h_i = H_iG^{-1}g(\frac{\sigma_i^2}{0.5\sigma_a^2})$, where $\sigma_a^2$ is the total additive genetic variance. Note that once $G^{-1}g$ is calculated, then different matrices $H_i$ could be applied to predict the genetic effects of different loci (assuming that genotypes are available for each
locus and that each \( H_i \) can be constructed relatively easily). The ratio of \( \sigma_i^2 \) to \( 0.5\sigma_e^2 \) is also assumed to be known.

**Marker-QTL Relationship Matrix.** If the single locus was a major gene or a quantitative trait locus, then the infinitesimal model no longer applies. The procedures just described would only be an approximation that depends on the ratio of \( \sigma_i^2 \) to \( 0.5\sigma_e^2 \). The larger the ratio is, the poorer the approximation would be. Even so, this procedure may give a useful first indication of the magnitude of the QTL effects. These comments also apply to marker loci, and depend on how close the marker is to the QTL.

Now assume that the genotypes of animals are for a marker locus and not for a single QTL. The marker and QTL loci have a recombination rate between them of \( r \) units. Then the relationship matrix (\( M \)) for the QTL is constructed using the marker genotypes and the fact that there is recombination between the marker and QTL. There are several known marker alleles, but the number of QTL alleles is not known. Usually the assumption is made that there are two QTL alleles, but this is not required here. The construction of \( M \) follows the methods of van Arendonk et al. (1994) exactly. Note that this matrix has a non-zero determinant and can be inverted. Because the markers are linked to the QTL and recombinations occur each generation, the marker-QTL effects are expected to diminish with time (re-establishing linkage equilibrium).

**Application to reproduction traits in swine.** Information on several markers (Jiang et al., unpublished data) were available for identifying and estimating QTL effects on reproduction traits in pigs from crosses between two breeds, Meishan and European Large White. Records of number of piglets born alive on 176 sows (F2 generation) from crosses between sixteen Meishan (4 males and 12 females) and 13 European Large White (4 males and 9 females) were obtained (Archibald et al., unpublished data). Also, marker information (SLIT and FGG17d markers) was available for all animals. The F2 sows were born in 1991, 1994 and 1995. Gestation length varied between 108 and 119 days. This effect was assumed to be a classified factor from 112 to 116 days. Age was also classified into 10 groups. Variances were not estimated due to the low number of records. Instead the genetic variance was 0.73 and the environmental variance was 5.96, as indicated by Irgang et al. (1994). Breeding values for number of piglets born alive were estimated from the following model:

\[
y_{ijklm} = \text{year}_i + \text{gestation length}_j + \text{age}_k + \text{gamete1}_l + \text{gamete2}_l + \text{residual}_m
\]

Where \( y_{ijklm} \) is the number of piglets born alive of \( m \)th sow in year \( i \), gestation length \( j \), and age \( k \). Two gametes were only random effects in the model with the gametic relationship matrix \( G \). In order to estimate the QTL effect linked to these markers and also the relationship between the recombination rate and ratio of QTL to total genetic variance and the magnitude of QTL effect, 8 combinations (4 different recombination rates and 2 different ratios of QTL to total genetic variance) were used for each marker. The QTL effect linked to SLIT marker ranged from -0.02 to -0.82 for first allele and from 0.02 to 0.08 for second allele. The corresponding values for FGG17d marker ranged from -0.01 to -0.08 and from 0.01 to 0.03 for first and second allele, respectively. These values correspond to range of 0.04 - 0.17 and 0.02 -
0.07 piglet for SLIT and FGG17d markers, respectively, when first allele is substituted with the other.

RESULTS AND DISCUSSION

The method described in this paper, assumes that $G(0.5\sigma_v^2)$ is applicable to all loci affecting the trait of interest, and that the goal is to partition it into effects due to individual QTL or markers. Thus, predicted gametic effects for each animal are calculated from mixed model equations. The relationship matrix for the single locus is assumed to be just one of an infinite number of such relationship matrices that are part of $G$. Using the principles from Henderson (1985), single locus gametic effects can be estimated from the overall predicted gametic effects by relatively simple computations. This could be a useful way of screening many markers for their potential usefulness in selection. Once a potentially useful marker has been identified, then other models could be applied.

A more appropriate model is that of Fernando and Grossman (1989) in which the effects of the marker alleles and the remaining polygenic effects are estimated simultaneously. The model, in matrix notation, is

$$y = Xb + Zu + Wv + e,$$

where $u$ represents the vector of polygenic effects that are independent of the QTL with one effect per animal, $v$ is a vector of marker-QTL allelic effects (two for each animal), $b$ is a vector of fixed effects, and $X$, $Z$, and $W$ are incidence matrices relating observations to the corresponding effects. Also,

$$\text{Var} \begin{pmatrix} u \\ v \\ e \end{pmatrix} = \begin{pmatrix} A\sigma_u^2 & 0 & 0 \\ 0 & M\sigma_v^2 & 0 \\ 0 & 0 & 1\sigma_e^2 \end{pmatrix},$$

where $\sigma_u^2 = \sigma_v^2 + 2(\sigma_e^2)$.

The matrix $A$ is the same as in any animal model. As shown previously, $u$ can be replaced with $g$, a gametic effect (two per animal) and $A\sigma_u^2$ can be replaced with $G(0.5\sigma_u^2)$, and the design matrix $Z$ would have to be the same. The MME are

$$\begin{pmatrix} X'X & X'Z \\ Z'X & Z'Z + G^{-1}k_u \end{pmatrix} \begin{pmatrix} b \\ g \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \end{pmatrix},$$

where $k_u = 2\sigma_u^2/\sigma_v^2$, and $k_v = \sigma_v^2/\sigma_e^2$.

The procedure described in this paper depends on the difference between $[G(0.5\sigma_u^2) + M\sigma_v^2] - G(0.5\sigma_u^2)$.

The larger $\sigma_v^2$ is, the better the model of Fernando and Grossman (1989) would be. However, no sensitivity study has been performed to determine when the approximation method should not be used.
If a finite loci model is assumed, then $G(0.5\sigma_a^2)$ can not be used for the polygenic effects in the Fernando and Grossman (1989) model, and the approximation procedure should not be used either. Instead $G(\sigma_a^2)$ should be derived for each locus and summed over the number of loci affecting that trait. Assuming the combined matrix can be easily inverted, then the partitioning of effects as described in this paper can be used. That is, $H(0.5\sigma_a^2) = \sum_i M_i \sigma_i^2$, where $f$ is the number of loci affecting a trait, then the individual loci effects could be obtained from $\hat{h} = M_i H^{-1} \hat{g}(\frac{\sigma_i^2}{0.5\sigma_a^2})$ for $i = 1$ to $f$. However, most likely $H^{-1}$ would be difficult to compute for a large number of animals.

CONCLUSIONS
The described procedure in this paper is based on some assumptions that may not work under other circumstances. However, it is simple and may be useful for screening of potential markers particularly in a finite loci model.

REFERENCES