Chapter 17

Genetic evaluation at individual QTL

Brian Kinghorn and Julius van der Werf

University of New England

Evaluation of Genetic value or Breeding value?

Genetic value is the value of an animal's genes to itself. Breeding value is the value of an animal's genes to its progeny. In general, breeding value has been of much more importance to animal breeders - it reflects the merit that can be transmitted to the next generation. It is the sum of the average effects of alleles carried by the animal, and because of the large number of loci classically assumed, there is no power to capitalize on anything but the average effects of these alleles, as dominance deviations in progeny cannot be predicted under normal circumstances.

However, when dealing with individual QTL we have the power to set up matings designed to exploit favorable non-additive interaction in the progeny

Approaches to QTL evaluation

Five approaches to genetic evaluation using markers can be identified:

1. Marker association with merit across families. This relies on population level linkage disequilibrium. Other methods are generally better if pedigree information is available.
2. Within-family analysis, making inference about sires' QTL heterozygosity and marker-QTL linkage phases, in a framework similar to one used for marker assisted QTL detection. This leads to information for selection between sibs.
3. Use of markers to modify transmission probabilities in segregation analysis to calculate QTL genotype probabilities. Typically two QTL alleles are involved and QTL effects treated as fixed. This is probably preferable where few effectively distinct alleles are known to be segregating, and where dominance and/or epistasis are important.
4. Use of markers to infer probability of identity by descent of contributing QTL alleles, with QTL effect treated as random and no assumption about number of alleles at each QTL. This effectively extends 2. above to use all pedigree information and give QTL EBV's.
5. Use of genetic markers located within target QTL. This removes the need for trait measurements and pedigree information to evaluate animals at QTL of known effect, leading to Genotype Assisted Selection However, multiple alleleism means that only complete sequence markers are fully reliable, as otherwise QTL alleles of identical marker type can have different effects.

Analysis methods:

An analysis method which targets inference making at one or more known QTL segregating in the population has a number of ideal features, including:

- It should make *appropriate* use of all available information. This may involve inference about which meioses involve crossing over between the QTL and one or more marker loci.

- It should be able to make inference about genotype of individuals at the QTL, possibly by calculation of probabilities for all genotype states.
- It should be able to estimate QTL genotype effects and allele frequencies
- It should be computationally feasible, especially for multiple QTL and many marker loci.
- It should account for 'genetic background effects' and have a term for residual polygenic effects.

An iterative sampling approach, such as Gibbs sampling, will cover the first three above, as it samples across the range of possible values and states, including all possible patterns of realised QTL- marker crossing over in the population. Results are generally unbiased, although there can be problems with unusual distributions of effects or states, for example when genotype probabilities are sampled at 0 or 1 for small groups of related animals, and the sampling chain becomes 'stuck' at these values. Unfortunately, these approaches are generally slow.

The next pages describe analytical analysis methods which are faster, and relate to approaches 3 and 4 above.

Modeling effects at the QTL genotype

The information on genetic markers can be used to make an inference about a likely genotype at a single locus. We know more about the segregation of alleles at this locus compared to the effects of alleles at all other (unknown) loci. We distinguish between effect of QTL vs. polygenic effect, the last representing the allelic effects at unknown loci. The model for the observations is

$$y = Xb + Wq + Zu + e$$

with fixed effect in *b*, qtl effects in *q* and polygenic effects in *u*. The design matrix W relates observations to QTL genotype. The first and second moments of the model need to be specified,

Expectation: $E(y) = Xb + E(Wq)$

Variance: $\text{var}(y) = \text{var}(Wq) + Z\,\text{var}(u)Z' + \text{var}(e)$

In the polygenic mixed model, breeding values are random effects, and
$\text{var}(u) = A\sigma_a^2$ where σ_a^2 is the (polygenic) additive (poly) genetic variance.

A model with both QTL and polygenic effects is indicated as *Mixed Inheritance Model*, referring to the fact that inheritance at QTL not necessarily the same as of polygenic effects. For example, the genotype at the QTL of two animals (e.g. parent and offspring) can be exactly the same for the QTL, whereas for polygenic effects the covariance is at most 50% of the total polygenic variance.

An accurate model ensures correct estimation of all relevant effects. Hence, polygenic effects are estimated while accounting for effects at the QTL, whereas QTL effects are estimated with an account for polygenic variation. Notice that this model is general and can therefore be used as a more general approach for QTL detection, independent of design.

QTL-genotype as a fixed effect:

These methods are based on a 2-step iterative scheme of, firstly, calculating QTL genotype probabilities using segregation analysis, and secondly, regressing phenotypes on these probabilities (Kinghorn et al., 1993; see figure below) or carrying out regression weighted by these probabilities (Meuwissen and Goddard, 1997). In both cases, fixed effects and polygenic breeding values are also fitted.

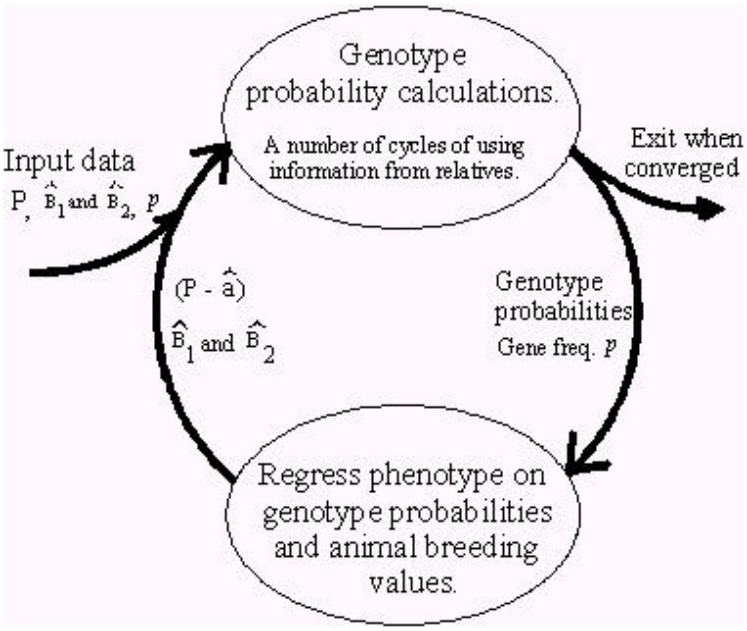


Figure. See Chapter 3 for more detail.

This is a useful approach when there is a limited number of genotypic effects at the QTL, i.e. there is a limited number of alleles, and the effect of different genotypes are equal across families. As the model estimates genetic effects of all genotypes, it can easily accommodate dominance at the QTL (and epistasis if effects at more QTL's were involved). The design matrix W relates observations to QTL genotype. This would be a regular design matrix if QTL genotype is known. However, we have generally only knowledge of *genotype probabilities* at the QTL. The matrix W can contain such probabilities. Genotype probabilities can be calculated with segregation analysis, and information on markers can be used to modify transmission probabilities in segregation analysis to calculate QTL genotype probabilities. A general method may again be applied, e.g. some animals may have marker information, but others don't.

With QTL genotype as fixed effect the model specifications for the QTL effect are E(Wq) = Wq and
var(Wq)=0.

Marker information is accommodated by modifying transmission probabilities at the segregation analysis step, according to prevailing marker genotypes (Meuwissen and Goddard, 1997).

$$\begin{pmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\alpha_u & Z'W \\ W'X & 0 & D \end{pmatrix} \begin{pmatrix} b \\ u \\ q \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \\ W'y \end{pmatrix} - \begin{pmatrix} r \\ 0 \\ 0 \end{pmatrix}$$

$$D = \begin{pmatrix} \sum_i W_{i1} & 0 & 0 \\ 0 & \sum_i W_{i2} & 0 \\ 0 & 0 & \sum_i W_{i3} \end{pmatrix} \qquad r = \begin{pmatrix} \sum_i W_{i1}\hat{u}_j(1) \\ \sum_i W_{i2}\hat{u}_j(2) \\ \sum_i W_{i3}\hat{u}_j(3) \end{pmatrix}$$

Note that D is a summation of genotype probabilities rather than a crossproduct of the coefficients in W. This reflects that the Meuwissen and Goddard method uses a summation of three regressions on known genotype, each weighted by the probability of having that genotype. This method should behave better (possibly showing showed less bias) than the method of Kinghorn et al. (1993), which is essentially a joint regression of phenotype on fixed and random effects and genotype probabilities at the QTL.

QTL genotype as random effect

Use of markers to infer probability of identity by descent of contributing QTL alleles, and no assumption about number of alleles at each QTL. QTL effects are estimated to be different for different families.

This model was originally proposed by Fernando and Grossman (1989). The effect of a QTL is modeled as the sum of the two gametic effects:

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

The vector v is twice the number of animals in the analysis. For each animal it contains a paternal and a maternal gametic effect.

The variances of the random effects are

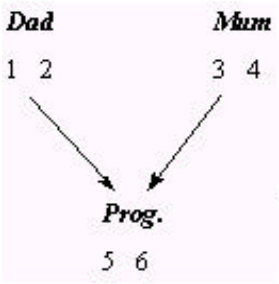
$$\text{var}\begin{pmatrix} u \\ v \\ e \end{pmatrix} = \begin{pmatrix} A\sigma_u^2 & 0 & 0 \\ 0 & G_{v/r}\sigma_v^2 & 0 \\ 0 & 0 & I\sigma_e^2 \end{pmatrix}$$

Where G is the gametic relationship matrix. It contains the probability of 'identity' between each of the two alleles in each individual at the QTL and these probabilities depend on marker information, and the assumed recombination rate between marker and QTL (see next). The subscript r indicates that for a given data set calculation of G depends on the assumed r. And the mixed model equations are

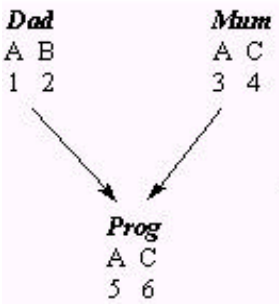
$$\begin{pmatrix} X'X & X'Z & X'W \\ Z'X & Z'Z + A^{-1}\alpha_u & Z'W \\ W'X & W'Z & W'W + G_{v/r}^{-1}\alpha_v \end{pmatrix} \begin{pmatrix} b \\ u \\ v \end{pmatrix} = \begin{pmatrix} X'y \\ Z'y \\ W'y \end{pmatrix}$$

This approach to genetic evaluation at marked QTL calculates probabilities of identity by descent of QTL alleles between gametes. These can be used to calculate animal EBV's at the QTL, much as we use coefficients of relationship to estimate 'polygenic' breeding values.

Marker information is used to calculate probability of identity by descent for alleles in different individuals (Goddard, 1992) or in different gametes (Fernando and Grossman, 1989; van Arendonk et al., 1994, Wang et al., 1995). This approach is most properly dealt with at the gametic level. The figure gives a simple illustration of how marker genotypes can help to more accurately build a gametic relationship matrix.



		Dad		Mum		Prog.	
	Site	1	2	3	4	5	6
Dad	1	1	0	0	0	.5	0
	2	0	1	0	0	.5	0
Mum	3	0	0	1	0	0	.5
	4	0	0	0	1	0	.5
Prog	5	.5	.5	0	0	1	0
	6	0	0	.5	.5	0	1



		Dad		Mum		Prog.	
	Site	1	2	3	4	5	6
Dad	1	1	0	0	0	.9	0
	2	0	1	0	0	.1	0
Mum	3	0	0	1	0	0	.1
	4	0	0	0	1	0	.9
Prog	5	.9	.1	0	0	1	0
	6	0	0	.1	.9	0	1

Figure. Gametic relationship matrices (GRM) for a QTL, on the right, are of dimension 6 sites x 6 sites for the simple 3-animal pedigree shown. Elements of the GRM are probability of identity by descent of the alleles at the prevailing pair of sites. In the upper GRM, no marker information is available, and, for example, probability of identity by descent between sites 4 and 6 is 0.5, as site 6 (maternal) could have inherited from sites 3 or 4 with equal probability. In the lower GRM, a marker with alleles A, B and C is available, and for example, probability of identity by descent between sites 4 and 6 is 1, for the marker locus. If the QTL is linked with a recombination fraction of 0.1, then the probability of identity by descent between sites 4 and 6 is 0.9, for the QTL, with a 0.1 probability

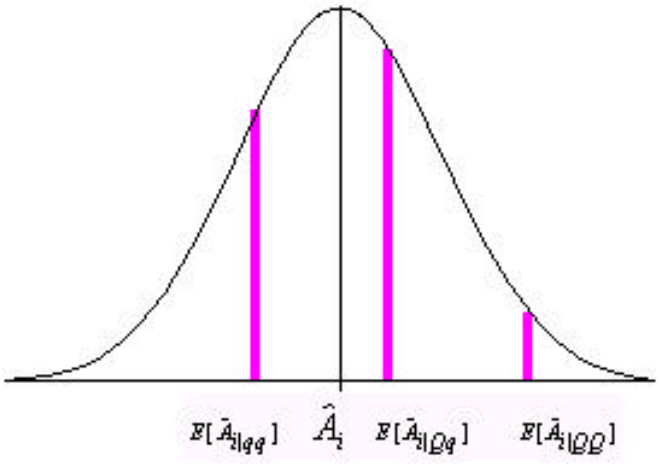
(in the event of recombination) for sites 3 and 6. Special attention is required where there is ambiguity of marker allele inheritance (Wang et al., 1995).

For any one pedigree there is a realized gametic relationship matrix, and this contains only 1's or 0's as elements, simply because each pair of QTL alleles are either fully identical by descent, or not at all. This concept has been used to test different methods for building predicted gametic relationship matrices through simulation of replicated populations and averaging of realised gametic relationship matrices (B.E. Clarke, unpublished). Different methods of building the GRM give slightly different results, and the differences in resulting EBV's are very small indeed.

One weakness of this approach to evaluating animal for a QTL is that no inference is made about QTL genotype of individuals, or, it is not so easy to accommodate dominace effects at the QTL.
This can be important where there are known non-additive effects involved (as mentioned above) , as it leads to possibilities for mating structures more effectively. However, an approach to getting the required QTL genotype probabilities from their estimated breeding values is given on the next section.

Genotype Probabilities from the GRM method

Having used the gametic relationship matrix approach to estimate breeding values at a QTL, there is an approach to calculating genotype probability is for each QTL genotype for each individual (Kinghorn and Clarke, 1996). For each animal, the expectation of its estimated breeding value is calculated conditional on each possible QTL genotype. These are then related to the actual estimated breeding value ($\hat{A}_i$) and its error distribution as the figure. The heights of this distribution at each QTL genotype are proportional to the genotype probabilities for this animal.



Heights of the distribution of $\hat{A}_i$ *(EBV at QTL for animal i) at the expectations conditional on QTL genotype are proportional to genotype probabilities.*

QTL allele frequencies can then be estimated by use of a simple counting procedure using genotype probabilities, with iteration to achieve convergence. Given QTL effects and genotype probabilities, variance due to the individual QTL can be simply calculated, such that QTL breeding values can be estimated based on relevant priors. This helps to overcome one weakness in the identity-by-descent method - but without further extension, the assumption of known QTL genotype effects remains.

Dissecting the genotype

In this lecture models were presented with effects of single QTL. In the near future, the number of QTL's targeted will probably increase, and the polygenic component will slowly be replaced by multiple QTL effects, the inheritance of each of them followed by marker brackets or more generally by information on haplotypes. Nejati-Javaremi et al. (1997) presented the concept of total allelic relationship, where the covariance between two individuals was derived from allelic identity by descend, or by state (based on molecular marker information), with each location weighted by the variance explained by that region. This in contrast to average relationships used in the numerator relationship matrix, which is derived from pedigree. Nejati-Javaremi et al. (1997) showed that using total allelic relationship resulted in higher selection response than pedigreed based relationships, because it more accurately accounts for the variation in the additive genetic relationships among individuals. For example, two full sibs have on average 50% of their alleles in common, but at specific loci they may either have none, or all alleles in common. Therefore, the gain of following inheritance at specific genome locations contributes to more accurate genetic evaluation, and is able to more specifically deal with within and between loci interactions and specific modes of inheritance at different QTL .

References

Fernando, R.L.; Grossman, M. (1989) Marker assisted selection using best linear unbiased prediction. *Genetics, Selection,.Evolution* **21**, 467-477.

Goddard, M.E. (1992) A mixed model for analyses of data on multiple genetic markers. *Theoretical and Applied Genetics* **83**, 878-886.

Kinghorn, B.P.; Kennedy, B.W.; Smith, C. (1993) A method of screening for genes of major effect. *Genetics* **134**, 351-360.

Kinghorn, B.P. and Clarke B. E. 1997. Genetic evaluation at individual QTL. Animal Biotechnology, 8:63-68.

Meuwissen, T.H.E. and Goddard M.E. 1997. Estimation of effects of quantitative trait loci in large complex pedigrees. Genetics 146:409-416.

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

Van Arendonk, J.A.M.; Tier, B.; Kinghorn, B.P. (1994) Use of multiple genetic markers in prediction of breeding values. *Genetics* **137**(1), 319-329.

Wang, T., Fernando, R.L., Van der Beek, S, Grossman, M, and van Arendonk, J.A.M. 1995. Covariance between relatives for a marked quantitative trait locus. Genet. Sel. Evol. 27:251-274.