Research in QTL Analysis

Identity-by-Descent (IBD) and genotype probabilities

To maximise the information obtained from genotyping animals for both linked and direct markers, improved methods for estimating IBD and genotype probabilities are required. Most existing methods are either infeasible for the large complex pedigrees common in livestock, or produce biased results in complex pedigrees (those with marriage or inbreeding loops), or require starting values that are close to optimum to guarantee convergence to regions of high likelihood.

Researchers at AGBU have developed a new algorithm that directly samples the parameter space of genotypes. A descent graph, describing the flow of genes in the population and consistent with the pedigree and genotypic data is drawn for each sample. It is used to determine the (ordered) genotype of each individual. The method is known as 'conditional sampling'. The mean of these independent samples is used as an estimate of the probabilities, with the mean of the descent graphs estimating the IBD probabilities and the mean of the genotypes estimating the genotype probabilities. The algorithm is suitable for large, complex pedigrees, and through the use of weighting, both IBD and genotypic probabilities are estimated for all individuals. Work continues on faster methods for including linked loci, and on methods for dealing with conflicting genotype and pedigree data.

Which Animals Type to Genotype

Genotype Probabilities

It would be very expensive to genotype all animals in a flock or herd. Luckily, it is not necessary. Where the genotypes of some animals are known, the genotypes of many other animals in the pedigree can be determined exactly. For the others, the probabilities of all possible genotypes can be determined.

These results can be used to identify from the available individuals those which, if genotyped, would provide the most information to more precisely determine the genotypes of (other individuals of interest in) the population. Many possible objectives can be met. These include a limit on expenditure or a desire to know the genotypes individuals which are not available for direct genotyping e.g. a long-dead ancestor, an expensive AI sire. Sensitivity analysis is also available to indicate how useful some additional expenditure might be or conversely how funds could be conserved by genotyping fewer individuals.

The end result is a list of individuals to genotype. This result maximises information at minimum cost. The strategy makes full use of any genotypes already known, either in the herd or on relatives in the public domain. In deriving the list, consideration of all possible outcomes is made.

Multi-stage Genotyping

If time permits, a multi-stage genotyping strategy can be more efficient. In this case, a series of analyses are completed, one for each stage. The results from each stage of genotyping are used in determining the next stage.

By allowing a series of stages the genotypes of key individuals can be determined first. Their results will sharpen the genotype probabilities

of other ungenotyped individuals, thus reducing the requirement to genotype more individuals. The pricing structure of the test and other costs can be accommodated in these analyses.

Identity-by-Descent (IBD) Probabilities

The Gene Detective uses IBD probabilities for markers linked to QTL. The best animals to genotype to determine IBD probabilities may be different to the best animals to genotype to determine genotypic probabilities. The objective of minimising genotyping costs while maximising information on IBD is also available.

Finite Locus Models

The majority of models used in genetic evaluation systems are based on the infinitesimal model, which assumes the additive genetic effect is controlled through the summing of a large number of small and independent gene effects. While BLUP and REML are extremely successful in the prediction of breeding values and estimation of variance components, these statistical tools do have some limitations.

The prediction of specific combining ability and its exploitation, whether at the individual, breed or species level, is an important issue in breeding. The accurate prediction of dominance effects and possibly epistatic effects is highly desirable. On a within population level, the prediction of dominance variance is estimated using the theory that the genetic covariance observed among full-sibs is a quarter of the dominance variance. The variance can be estimated simply by including a family effect in the model. If more than two generations of pedigreed data exist and inbreeding is present, the estimation process becomes more complicated. Genetic evaluation in multi-breed populations is even more complex and the definition of the correct covariance structure is practically unworkable.

In what are termed "finite locus" models an individual's genetic value is modelled as the cumulative value of a series of genes, with each gene having a defined effect and frequency. Computation of variances and other parameters becomes more tractable as they can be computed empirically. A feature of our implementation of the finite locus model has been the use of the descent graph sampling algorithm, used in 'The Gene Detective', to sample genotypes at each locus. This algorithm is perfectly suited to finite locus modelling and will prevent any bias due to incomplete sampling of the parameter space. Bias stemming from incomplete sampling of the parameter space can be incorrectly construed as stemming from non-use of the appropriate prior distribution for gene effects. Our research has shown that choice of priors for gene effects has little bearing on the accuracy of the results. A much more critical factor is the issue of sampling the whole parameter space for genotypic configurations.

In a paper presented to the 13th AAABG a finite locus model was used to estimate genetic parameters from unselected, simulated populations. One hundred unlinked loci were simulated in order to generate a true polygenic trait. Predictions of genetic effects under a finite locus model were comparable to REML and BLUP methods in analysing the same data.

In collaboration with the Queensland Forestry Research Institute, two subtropical pines and their inter-specific hybrid were analysed using a finite locus model. Estimates for genetic parameters (which included the correlation between purebred and hybrid performance) were shown to be slightly higher and more precise to those estimated by ASREML.

AGBU
ANIMAL GENETICS AND BREEDING UNIT
A joint unit of NSW DPI and UNE