Breeding program design including MAS

THE EFFECT OF THE INCLUSION OF PEDIGREE DATA ON ESTIMATES OF CARRIER STATUS AT THE AGOUTI LOCUS IN SHEEP

J. M.Henshall[1], J. McNally[1] and B.J. Norris[2]

[1]CSIRO Livestock Industries, FD McMaster Laboratory, Armidale, NSW 2350
[2]CSIRO Livestock Industries, St Lucia, Queensland 4067

SUMMARY

A simulation study was conducted to examine the effect of including pedigree data when estimating genotype probability at the agouti locus. The effect of errors in the pedigree was also examined. The proportion of non-carrier progeny identified increased by as much as 20% when pedigree information was included, particularly if genotype estimates for maternal grandsires were available. Pedigree errors had to be at very high rates before they adversely affected the accuracy of genotype estimates.

INTRODUCTION

Approximately 0.5% of Merino lambs born each year are recessive black, and therefore around 13% of the Australian Merino flock are estimated to be carriers of the black pigmentation allele. Variation in the Agouti region is known to be responsible for the recessive black condition (Parsons, Fleet and Cooper 1999), with the causative allele identified as having 1 copy of agouti signalling protein coding sequence, while dominant alleles responsible for no pigmentation in the fleece are composed of between 2 and 5 tandem copies of the gene (Norris and Whan 2008). An estimate of the number of copies can be obtained using a DNA based test where the 'junction point' between each copy of the gene is amplified (Norris and Whan 2008). The assay (referred to here as the "agouti assay") is an asymmetrical competitive PCR where a fluorescently-labelled common reverse primer is added at 100 fold less then the specific forward primers allowing quantitative measurement of the total gene copy number. In a diploid animal the total gene copy number does not distinguish how many copies of the gene are on each chromosome. Thus, the distinction between white animals with multiple copies of the gene at both loci and carrier animals that are heterozygous for the single copy allele (recessive black allele) and a multiple copy allele, cannot be determined. Further, the quantitative nature of the assay produces probabilities for each gene copy number rather than calls made with certainty. The assay returns a real number with expected value $\frac{n-2}{n}$, where n is the total gene copy number. This series converges as copy number increases, so high copy counts in particular are not known with certainty.

The utility of the assay can be enhanced by taking account the constraints imposed by Mendelian transmission of copy number alleles in pedigreed populations. Assay values can be used to specify penetrance values relating to the unobserved genotype, and segregation analysis on pedigreed populations can then provide estimates of the probability of each underlying genotype configuration. For commercial sheep populations a half-sib data structure will be usual, but it is unlikely that the pedigree will be without error as, unless paternal and maternal DNA parentage is carried out, pedigree errors will occur on the dam side even in flocks that lamb in sire paddocks. In this paper we examine the benefit that can be gained by including pedigree in addition to sire in analysing assay results. As this increases the risk that incorrect information is included in the analyses we also consider the impact of pedigree errors on the estimated genotype probabilities.

MATERIALS AND METHODS

Simulation. The population was modelled on 10 small stud flocks each joining 5 rams to 200 ewes each year over 20 years. Of the 5 rams, 4 were home bred and 1 came from a linked flock. Selection was on a trait uncorrelated to fleece colour except that no black animals were selected in the final generation. Allele frequencies estimated from a Merino sheep population were simulated in the base population. In the final generation, assay values were simulated for 16 progeny from each sire and for the sires themselves. Variance around the expected assay value was simulated with a coefficient of variation (CV) of 3%. The simulation was repeated 20 times producing a total of 1,000 halfsib families for analysis. For each simulated dataset, pedigree errors were superimposed as mismothering events where pairs of lambs were assigned to each other's mother at rates of 0%, 1%, 2%, 4%, 8% or 16%. The mismothering events were within halfsib groups, such as might occur when lambing takes place in sire paddocks or when paternity is determined through a DNA test. No paternity errors were simulated as when performing the agouti assay it is only a small amount of additional work to do sire pedigree verification using markers, thus the sire can be assumed to be known without error.

Analysis. Initial analyses took place using the software package Mendel (Lange *et al*. 2001), which computes the full likelihood for complex pedigrees. This restricts the size of pedigrees that can be analysed. Other software packages are available that lift the restriction by applying approximations to the full likelihood, however these would have required modification to allow for the penetrance function derived from assay data. To explore the effect of including pedigree information (with errors) in analyses for larger pedigrees, we used halfsib family analyses with modified penetrance function values for the un-assayed dams to simulate information flow from relatives of the dams. Our motivation for this approach was as follows: when analysing halfsib families, with the sire and progeny assayed, sire genotype probabilities are estimated with high accuracy. Therefore, when the analysis includes pedigree information above the halfsib family the sire estimates will not change much, but information will flow through the dam to the progeny. Most of this information will come from the maternal grandsire, whose own genotype may be known with high accuracy. This can be modelled in a halfsib analysis by modifying the dam penetrance values to account for knowledge of her sire's genotype. We considered levels of certainty of the maternal grandsire's genotype of 100%, 50% and 0%. In all of the analyses the penetrance function was calculated assuming a CV of 3% for the assay values.

RESULTS AND DISCUSSION

Analyses using the complete likelihood. Using Mendel, we had difficulty in estimating genotype probabilities in data with pedigree errors. In many cases the software execution aborted as the problem was too complex. This may be due to very small likelihoods, exacerbated by the pedigree errors. Even without pedigree errors we were unable to reliably analyse the 5 halfsib families from a flock with more than 4 progeny per family and with more than grandparents linking the families. Results for non-carrier progeny analysed using 4 progeny per family, the most we could achieve using Mendel, are presented in Table 1. There was a small benefit to including the pedigree information, but it was less than the benefit gained by assaying 8 progeny per family and analysing the resultant data as a halfsib family.

Halfsib analyses using modified dam penetrance values. Probabilities for non-carrier progeny are presented in Table 2. A maternal grandsire genotype certainty of 0.0 is equivalent to performing a half-sib analysis, and mismothering is irrelevant as only sire data is used in the

analysis. The effect of including information about the genotype of the maternal grandsires is shown by the difference between the results with a maternal grandsire genotype certainty of zero and a non-zero maternal grandsire genotype certainty. Including maternal grandsire genotype data increased the precision of the assay, lifting the proportion of lambs declared non-carriers with 99.9% certainty from 61% to as much as 73%. For a flock that has applied the assay for more than one generation the genotype of the maternal grandsires may well be known with high accuracy, and it is clear that there is great benefit from including their data in the analysis. Mismothering has a negligible effect on progeny genotype probabilities. Even at very high rates (16%) the effect is barely noticeable. Estimated carrier probabilities for non-carrier sires and dams were similarly improved by the inclusion of information about the paternal grandsire's genotype, and the effect of pedigree errors was minor (results not shown).

Table 1. Proportions of non-carrier progeny achieving <0.1%, <1%, <5%, >95%, >99% and >99.9% estimated probability of being a carrier for the undesirable allele. No pedigree errors were simulated

Analysis method	Family size	Estimated probability of being a carrier					
		P<0.001	P<0.01	P<0.05	P>0.95	P>0.99	P>0.999
Pedigree	4	0.39	0.65	0.78	0.00	0.00	0.00
Halfsib	4	0.38	0.62	0.76	0.00	0.00	0.00
Halfsib	8	0.50	0.71	0.81	0.00	0.00	0.00

Table 2. Proportions of non-carrier progeny achieving <0.1%, <1%, <5%, >95%, >99% and >99.9% estimated probability of being a carrier for the undesirable allele. There were 16 progeny in each halfsib family, with progeny and sires assayed

Grandsire genotype certainty	Mismothering (%)	Estimated probability of being a carrier					
		P<0.001	P<0.01	P<0.05	P>0.95	P>0.99	P>0.999
0.0	NA	0.61	0.74	0.81	0.00	0.00	0.00
0.5	0	0.67	0.77	0.82	0.00	0.00	0.00
0.5	1	0.67	0.77	0.82	0.00	0.00	0.00
0.5	2	0.67	0.77	0.82	0.00	0.00	0.00
0.5	4	0.67	0.77	0.82	0.00	0.00	0.00
0.5	8	0.67	0.77	0.82	0.00	0.00	0.00
0.5	16	0.66	0.77	0.82	0.00	0.00	0.00
1.0	0	0.73	0.79	0.85	0.00	0.00	0.00
1.0	1	0.73	0.79	0.85	0.00	0.00	0.00
1.0	2	0.73	0.79	0.85	0.00	0.00	0.00
1.0	4	0.73	0.79	0.85	0.00	0.00	0.00
1.0	8	0.73	0.79	0.85	0.00	0.00	0.00
1.0	16	0.72	0.79	0.85	0.00	0.00	0.00

For carrier progeny, estimates of probabilities of being a carrier were improved by the inclusion of information about the maternal grandsire's genotype, and again the effect of pedigree errors was minor (Table 3). Most importantly, the probability of being declared a non-carrier was

not increased by the inclusion of maternal grandsire data, even in pedigrees with high rates of mismothering. For carrier sires and dams, inclusion of maternal grandsire genotype information had a similar effect on estimated probabilities of being a carrier (results not shown).

Table 3. Proportions of carrier progeny achieving <0.1%, <1%, <5%, >95%, >99% and >99.9% estimated probability of being a carrier for the undesirable allele. There were 16 progeny in each halfsib family, with progeny and sires assayed

Grandsire genotype certainty	Mismothering (%)	Estimated probability of being a carrier					
		P<0.001	P<0.01	P<0.05	P>0.95	P>0.99	P>0.999
0.0	NA	0.00	0.00	0.01	0.65	0.61	0.57
0.5	0	0.00	0.00	0.01	0.68	0.65	0.59
0.5	1	0.00	0.00	0.01	0.68	0.65	0.59
0.5	2	0.00	0.00	0.01	0.68	0.64	0.59
0.5	4	0.00	0.00	0.01	0.67	0.65	0.59
0.5	8	0.00	0.00	0.01	0.67	0.65	0.59
0.5	16	0.00	0.00	0.01	0.67	0.64	0.59
1.0	0	0.00	0.00	0.01	0.71	0.70	0.66
1.0	1	0.00	0.00	0.01	0.71	0.70	0.66
1.0	2	0.00	0.00	0.01	0.71	0.69	0.66
1.0	4	0.00	0.00	0.01	0.71	0.69	0.65
1.0	8	0.00	0.00	0.01	0.71	0.69	0.65
1.0	16	0.00	0.00	0.01	0.71	0.68	0.65

CONCLUSIONS

When using Mendel to analyse small families of 4 halfsibs, including pedigree information in the analysis produced a small increase in the precision of estimates of genotype probabilities. Analyses on larger families of 16 halfsibs using our simplified model, where pedigree information was simulated by including knowledge of the maternal grandsire's genotype, suggests that greater improvements are likely to be achieved where information on genotype probability is available for maternal grandsires. Although the estimation can be carried out without software for segregation analysis in deep pedigrees, ideally all of the available data would be analysed together. Pedigree errors on the dam side had negligible effect on accuracy of estimates.

REFERENCES

Lange K., Cantor R., Horvath S., Perola M., Sabatti C., Sinsheimer J. and Sobel E. (2001) *Amer. J. Hum. Genetics.* **69(Supplement)**: A1886.

Norris, B.J. and Whan, V.A. (2008) *Genome Research*, **18**:1282.

Parsons, Y.M., Fleet, M.R. and Cooper, D.W. (1999) *Aust. J. Agric. Res*. **50**:1099.