Proceedings, 10th World Congress of Genetics Applied to Livestock Production

# Evaluating Sires from Commercial Progeny Data Using Pooled DNA

**A. Bell[1], J. Henshall[1], R. McCulloch[2] and J. Kijas[2]**
[1]CSIRO Animal, Health and Food Sciences, Armidale [2]CSIRO Animal, Health and Food Sciences, Brisbane

**ABSTRACT:** This study proposed to use genotypes from pooled DNA samples of progeny to enable evaluation of sires' performance in a commercial flock, for a trait that is difficult to measure in the stud environment. Blood samples were collected from 786 one year old Merino sheep in a commercial flock in southern Victoria, Australia. Animals were grouped into cohorts by dag score phenotype and sex (male or female). Eighty blood samples were collected within each cohort and randomly allocated to two equally sized pools. Pooled samples were assayed with the Ovine SNP 50 chip and allele frequencies estimated for each SNP in each pool. Blood samples were also collected from 33 sires of the commercial flock and genotyped. The contribution of sire alleles to each pool was estimated and these contributions were not randomly distributed across pool dag scores (P = 0.015). Genotyping pooled DNA of progeny and genotyping sires can be utilized as an estimate of sire performance for a difficult to measure, but commercially important trait.
**Keywords:** sheep; genotype; phenotype; pooled DNA

## Introduction

In the Australian sheep industry most genetic improvement is undertaken in the stud sector. Mostly this limits genetic progress to those traits which exhibit variability in the stud environment. Ram buying clients of studs can be in different geographical locations to the stud, in a range of environments, and traits of interest may be expressed to a much greater extent. This presents the opportunity to measure and estimate the sires' contribution by progeny testing on commercial properties using sires from identified studs. Phenotypes of progeny in a commercial environment, where there are typically very large numbers within a flock run as one contemporary group from birth, can be utilized if genotype information can be sourced cost effectively, and the sires' genotypes can be used to estimate the genetic contribution of the sires to the measured phenotypes.

Commercial sheep producers may source all their rams from one particular stud and may have done so for several years. Consequently, favorable pedigree linkages may exist from both the paternal and maternal (grand sires for example) sides of the pedigree. Pedigree linkages from the sires used in the commercial flock back to their home stud can be strategically used to exploit the range of traits under selection. One trait of particular interest to a commercial sheep breeder in southern Victoria is soiling of the breech area with faecal material (dags). Dag formation is caused by the adhesion of faecal material to the wool in the breech area and is measured on a scale of 1-5.

Breech soiling of Merino sheep devalues the wool harvested, increases the time commitment and cost of management practices, and predisposes the animal to an increased risk of flystrike. Estimates of heritability for dag score (DS) can range from 0.1 up to 0.25 depending on the age of the animal at testing (Pollott et al, 2004). DS is a relatively easy trait to score. The scoring is undertaken by visual assessment and does not requiring restraint of the animal nor the analysis of a biological sample at another location. A large number of animals (at least 800) can be phenotyped in a day, as part of a routine management event such as drenching with anthelmintic. The ease of the phenotyping technique allows it to be a candidate for high throughput phenotyping, and it can be readily combined with bleeding the phenotyped animals to obtain a blood sample for DNA.

Recent advances in genomics, in particular the availability of high throughput SNP assays, offer new opportunities for progeny testing. Commercial flocks can contain many thousands of individuals, in large contemporary groups. Multiple sire joining is practiced so individuals are not pedigreed, and individual DNA parentage could be cost prohibitive. With SNP assays on pooled blood samples there is not a need for individual pedigree or performing SNP assays on all individual progeny (Craig et al, 2009). Flocks can also be sampled opportunistically, at times when variation for DS exists.

The aim of this project was to estimate sire contributions to dag score phenotypes using pooled DNA from progeny in a commercial environment. Using this technique data from commercial farms could be routinely used to increase the rate of genetic progress in studs for a range of traits.

## Materials and Methods

The commercial Merino producer that collaborated in this study is located in southern Victoria, an area with a high incidence of dags, and rams are sourced from the Monaro region of southern NSW. There is currently no opportunity for this stud to select for reduced dags, as the low incidence of dags in the Monaro precludes observable variation in DS.

**Animals.** The 2600 Merino sheep available for participation in this study were born in Spring 2011 and

were 13-14 months of age at the time of phenotyping. The sheep were managed in two contemporary groups based on sex, ewe (female) and wether (castrated male). Dag score phenotypes on a subset of each contemporary group were measured during routine scoring for breech soiling. Scores were based on industry standards (Visual Sheep Scores, 2007) with DS phenotypes ranging from 1 (no soiling) to 5 (heavy soiling). Scored sheep were identified by a colour coded clothes peg, indicating which score they had received. A maximum of 80 sheep within sex and DS phenotype cohort were identified and subsequently bled for a DNA sample. There were only 66 female sheep present with a DS of 5. A total of 786 blood samples were collected with 6 mL EDTA (K+) vacutainers from the identified sheep by jugular venupuncture. Blood samples were also collected from the 33 (out of 45 in total) sires of the 2011 progeny remaining on the property. All samples were stored on ice and shipped overnight to Brisbane for processing.

**DNA pooling and extraction.** Within each DS (1 to 5) by sex cohort, blood samples were allocated at random to two equally sized pools, giving 40 samples per pool, and 33 samples per pool of the DS 5 female cohort. Pooled blood samples were created by combining 20µL of fresh whole blood from each sample. 200µL of this pooled blood was used to extract genomic DNA following the manufacturer's instructions (Qiagen DNeasy® Blood and Tissue Kit).

Blood samples from sires were treated individually and DNA extracted using the same method.

**Genotyping.** Each sample (20 pooled + 33 sires) was assayed with the Ovine SNP 50 chip. The Illumina software provides discrete genotype calls and estimates of the B allele frequency for each SNP. In subsequent analyses we used the discrete genotype calls for the samples from sires and the estimated B allele frequencies for the pooled samples.

**Statistical analyses.** For each pool, we used a variant of the method of Kinghorn et al. (2010) to estimate the contribution from each sire. Essentially, this involves fitting a model $Y = X\beta + \varepsilon$ where $Y$ is a vector of allele frequencies for the pool, $X$ is a matrix of sire genotypes (or family genotypes in the case of Kinghorn et al. (2010)), $\beta$ is a vector of sire contributions to pools and ε is an error term. This model was developed to estimate family contributions to pools, and the vector $\beta$ is constrained to have elements greater than or equal to zero, and to sum to 1.0. There are several deficiencies in using this approach for our data: we were estimating sire contributions, not family contributions, we had DNA from only 33 of the 45 sires contributing genes to the pools, and we had no DNA at all from the dams. Despite this we used the constrained least squares pcls() procedure in the mgcv package (Wood (2000)) in the statistical software R (R Core Team, (2013)) to fit the model. We applied the constraint that contributions were greater than or equal to zero, but unlike in Kinghorn et al. (2010), we did not apply a constraint to ensure that the contributions summed to one. Instead, contributions were scaled by dividing by the sum. Little scaling was required, as the range of summed contributions produced by pcls() was from 0.992 to 1.034. As a quarter of the sires were not genotyped we are likely partially allocating the contributions for ungenotyped sires to closely related individuals amongst the 33 genotyped sires. The same applies for the contribution from the dams, and as the stud has been the sole supplier of rams to the commercial flock for some time the maternal grandsires of the progeny are likely to be related to the 33 sires. To determine whether the distribution of sire contributions to pools was influenced by DS, we conducted an analysis for each sire, regressing the vector of his pool contributions on the pool DS. Progeny sex was included in the model, and P values were confirmed through permutation testing. The permutation test involved 100 replicates on each sire with phenotypes within sex and within replicate permuted. P values were allocated to significance level classes ($P > 0.05$, $0.05 \geq P > 0.01$, $0.01 \geq P > 0.005$, $0.005 \geq P > 0.001$, $0.001 \geq P > 0.0005$ and $P \leq 0.0005$), and the distribution of counts for the real data compared to the distribution of counts for the permuted data using a Pearson's chi squared test with simulation used to derive the significance level (function chisq.test() in R (R Core Team (2013)).

## Results

**Sire contributions.** Estimates of sire contributions to pools ranged from zero to 0.11, and the distribution is plotted in Figure 1. The left hand panel (A) shows the gene contribution to all sire by pool (33 sires x 20 pools) contributions, and in the right hand panel (B) the distribution of sire mean contributions to pools is displayed. Four sires make very small contributions but no sires dominate; none contribute more than twice the expected 3.03%.

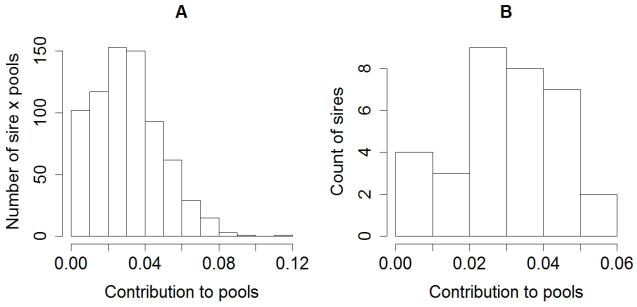


**Figure 1. Histograms of sire contributions to pools**.

**Associations with DS.** When the sire contributions are regressed on DS we found that far more sires were significant at various levels than would be expected by chance. As shown in Table 1, 30% (10 out of 33) sires were significant at the 5% level, far more than would be expected by chance, and this applied also to significance levels of 0.01 (4 sires), 0.005 (2 sires) and 0.0005 (1 sire). Just in case these departures were due to a problem with our model we conducted a permutation test, and found no departures from expected proportions when the permuted data

were tested. When P values were allocated to significance level classes and the counts compared for the observed and permuted data, the chi square test was significant (P = 0.015), indicating that the variation observed between sires in the contributions to pools of different DS is likely to be real, as would be expected for a heritable trait.

**Table 1. Proportions of analyses that were significant at various levels for the observed data and for the data with phenotypes permuted. *0.0303 is one out of 33 sires.**

| | Significance level | | | | |
|---|---|---|---|---|---|
| | **0.05** | 0.01 | 0.005 | 0.001 | 0.0005 |
| **Observed** | 0.30 | 0.12 | 0.061 | 0.030* | 0.0303* |
| **Permuted** | 0.049 | 0.011 | 0.0048 | 0.0009 | 0.0006 |

Figure 2 illustrates the sire gene contribution estimates to dag score pool for the best sire of those genotyped, and the worst sire. As demonstrated, the sire identified as the best for DS (mean dag score of 2.56) contributes less genes as DS increases, and the sire identified as the worst contributes a higher proportion as DS increases (mean dag score of 4.34).

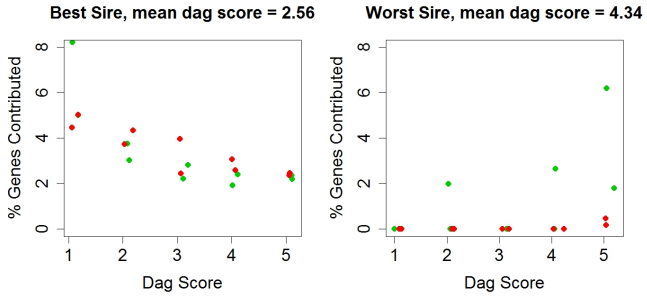


**Figure 2. Sire gene contributions (%) to dag score pools. Female progeny shown in green (2 pools per sex:phenotype), and male progeny shown in red. X-axis jittered to show over-plotted points.**

## Discussion

In this study we used SNP assays to link pools of blood from commercial progeny to blood samples from sires. There are various options for analyzing the data: SNP effects could be estimated in the commercial progeny and prediction equations applied to the sires, relationships between the pools and the sires could be estimated, or sire contributions to pools could be estimated. We used the last of these approaches, although our model was not ideal for our data as it did not account for the missing sires or the dam contribution to the progeny. Further work is required to determine the effect of the mismatch between the data and the model, and to develop models more appropriate to estimating contributions given missing data. To verify that our estimates of sire contributions to pools and sire mean DS are correct would require a parentage test on each of the 786 progeny, ideally with a very accurate parentage panel. We don't have these data, but the results we obtained appear to be consistent with what we would expect for a heritable trait.

Assuming that a validation was successful, the outcome of this study is the development of a technique using pooled DNA from progeny for evaluating sires in a commercial environment. Progeny tests are a valuable resource for improving genetic gain, and using this technique presents an opportunity to perform a progeny test in a situation where it was not previously plausible. A typical progeny test for this number of sires would have required the DNA pedigree testing of many hundreds of progeny, however by developing this technique, only 20 pools of DNA plus the 33 sires were genotyped. Data which can be used for selection decisions was obtained, and with the pedigree linkages back to the stud level, genetic improvement for this trait will be possible through estimated breeding values. The potential to use this technique for evaluating sires for a range of traits such as reproductive traits will be invaluable. Consideration would have to be given to the nature, size and number of progeny pools that would be constructed. Once the commercial sires have been genotyped, the progeny can be phenotyped for a number of traits, and further genotyping only conducted on a small number of pooled progeny samples.

## Conclusion

Low cost DNA tests can provide a link between phenotyped animals in a commercial flock and rams in stud flocks. Strategic and opportunistic sampling of commercial traits when abundant will allow the stud sector to expand the range of traits under selection.

## Acknowledgements

We thank the commercial property "Tramore" owners Fergus and Rosemary Irving for the use of their sheep and their assistance scoring the phenotypes, Brad Hine and Dominic Niemeyer for their proficient blood sampling for DNA, and Sonja Dominik for comments on the manuscript.

## Literature Cited

Craig, J. E., Hewitt, A. W., McMellon, A. E., et al (2009). Genome Res. 19:2075-2080

Kinghorn, B. P., Bastiaansen, J. W. M., Ciobanu, D. C. and van der Steen, H. A. M. (2010). Acta Agriculturae Scandinavica Section a-Animal Science. 60:3-12

Pollott, G. E., Karlsson, L. J., Eady, S. J. and Greeff, J. C. (2004). J. Anim. Sci. 82(10):2852-2864

R Core Team. (2013). R: A language and environment for statistical computing. Vienna, Austria: R Foundation for Statistical Computing.

Wood, S. N. (2000). Journal of the Royal Statistical Society Series B-Statistical Methodology. 62:413-428.

Visual Sheep Scores – Researcher Version, Australian Wool Innovation Ltd and Meat and Livestock Australia. November 2007. (Web Version 2- 2013 http://www.wool.com/Content/en-GB/Visual_Sheep_Scores_Researcher_Version_LR.pdf)