Proc. Assoc. Advmt. Anim. Breed. Genet. 20:278-281

SUCCESS RATES OF COMMERCIAL SNP BASED PARENTAGE ASSIGNMENT IN SHEEP

A. M. Bell[1,2], J.M. Henshall[1,2], S. Gill[2,3], K. Gore[2,4] and J.W. Kijas[5]

[1]CSIRO Animal, Health and Food Sciences, FD McMaster Laboratory, Armidale NSW 2350
[2]Sheep Cooperative Research Centre, Armidale NSW 2351
[3]Meat Livestock Australia, Sydney NSW 2059
[4]Animal Genetics and Breeding Unit, University of New England, Armidale NSW 2351
[5]CSIRO Animal, Health and Food Sciences, Queensland Bioscience Precinct, Brisbane QLD 4067

SUMMARY

Single nucleotide polymorphism (SNP) based parentage assignment is attractive as SNP are abundant in the sheep genome and amenable to high throughput and therefore lower cost genotyping. To examine the minimum number of SNP required to obtain high accuracy parentage assignment, blood samples were collected from 4 industry flocks and genotyping was undertaken. A maximum likelihood approach was applied to the genotypes to predict sire, dam and progeny within 3 of the sampled flocks, and dams within 1 sampled flock. A SNP based, flock specific methodology utilizing differing numbers and types of SNPs for estimating assignment rates was developed. Rates of assignment ranged from 99.5% to 77.7% across 3 flocks, with 0% incorrect assignments, with the exception of one panel in one flock for sire assignment, where the incorrect assignment rate was 0.1%. Rates of assignment varied from 62.2% to 28.3% with 0% incorrect assignments in the fourth flock, with the exception of one panel for dam assignment where the incorrect assignment rate was 0.1%, but only 60% of dams and 50% of sires within this flock were genotyped. Using 2 out of a potential 6 multiplexed panels of SNP markers gave high rates of correct paternity, but using 3 panels provided higher confidence and is recommended. This maximum likelihood approach using SNPs provides the basis for delivering highly accurate parentage determination for under AUD20, increasing the affordability of this as a powerful tool for industry.

INTRODUCTION

When parentage in a breeding enterprise is known, the rate of genetic progress can be improved by information from relatives when estimating breeding values by best linear unbiased prediction (BLUP EBVs) and accounting for maternal effects. Generating parentage records can be laborious and expensive due to the large amount of infrastructure required for artificial insemination programs, single sire mating strategies and mothering up or pedigree recording at lambing events. Inadvertent misallocation of lambs to dams can also occur particularly if dams are not scanned in lamb to obtain knowledge of the number of lambs expected and cross-mothering or mismothering occurs. DNA based methods of predicting parentage have been the focus of research in recent years, and the utilization of SNPs has reduced the cost of genotyping. Successful parentage testing requires a robust and technically accurate SNP genotyping platform coupled with a marker set containing SNP with high minor allele frequencies (MAF). The objective of this study was to develop an industry applicable low cost DNA based tool utilising SNP for determining sheep parentage.

MATERIALS AND METHODS

The SNP markers used in this project were identified by the International Sheep Genomics Consortium (ISGC). SNP were prioritised for use in parentage testing following analysis within a

spectrum of breeds using multiple genotyping platforms (Kijas *et al*. 2009; Kijas *et al*. 2012). In this experiment, 3 types of SNP were used to design sets or "multiplexes" of SNPs. A multiplex is a combination of SNP that can be assayed in a single reaction. The 3 SNP types were 1) ISGC parentage SNPs (those identified by the ISGC as suitable for parentage testing); 2) performance SNP (SNP identified as directly causing variation in phenotype or linked to mutations that cause phenotypic variation, for example, SNP linked to the Poll locus); and 3) filler SNP (SNP used to fill in around the first two types of SNP with a high MAF across a broad spectrum of breeds). A total of 383 SNPs were assigned to 6 multiplexed panels (named W1-W6). The number of markers in each multiplex ranged from 63 (W2, W3 and W4) to a maximum of 66 (W5). Details of each multiplex are shown in Table 1.

Table 1. SNP types within multiplexes

Panel	Total SNP	SNP Type		
		ISGC	Performance	Filler (MAF)
W1	64	38	6	20
W2	63	28	3	32
W3	63	18	1	44
W4	63	2	0	61
W5	66	0	0	66
W6	64	1	0	63
Total	383	87	10	286

Sheep Genetics (www.sheepgenetics.org.au) identified 4 industry representative flocks with varying levels of genetic relatedness between candidate sires, dams and progeny, and where knowledge of parentage was essential for breeding program purposes. Flocks 1, 3 and 4 are Merino enterprises, and Flock 2 comprises Dohne sheep. Blood cards were distributed by Sheep Genetics to the targeted producers. The blood cards were returned to Sheep Genetics, and sent to GeneSeek (USA) for DNA extraction and SNP genotyping using the SEQUENOM platform. Details of the flocks and number of sheep within each group in each flock are given in Table 2.

Table 2. Number of genotypes per flock and sheep group – details supplied by each flock owner

Flock	Sires		Dams		Lambs		Unknown	
	Total	Genotyped	Total	Genotyped	Total	Genotyped	Total	Genotyped
1	11	11	302	302	415	415	0	0
2	0	0	111	111	122	118	0	0
3	32	32	111	111	103	103	44	44
4	7	7	21	21	180	180	3	3

SNP data was only used if the sheep was known to be a sire, dam or lamb. Analysis of the genotype data was conducted without knowledge of the relationships between sires or the long term level of inbreeding. A maximum likelihood method was used for pedigree assignment (Marshall *et al.* 1998; Kalinowski *et al.* 2007; Kalinowski *et al.* 2010). Given the SNP data for a sire and a lamb, the likelihood that the sire is the parent is evaluated, along with the likelihood that the sire is not the parent of the lamb. The estimations use an assumed allele frequency for each SNP in the population, and an assumed genotyping error rate. As in Marshall *et al.* (1998) the logarithm (log) of the ratio (likelihood that the sire is the parent / likelihood that the sire is not the

parent) is referred to as the LOD score (from log odds). LOD scores were also estimated for lamb-dam pairs, and for lamb-sire-dam trios.

Simulation was used to derive an appropriate LOD threshold for each test type (lamb-sire pair, lamb-dam pair, or lamb-sire-dam trio) for each flock based on the observed allele frequencies. A total of 1000 progeny were simulated, each with randomly chosen sire and dam from the flock. Missing parents were simulated using allele frequencies estimated for the flock. For each simulated lamb, LOD scores were estimated for each sire and dam. For the most likely 5 sires and most likely 5 dams the LOD score was estimated for each of the 25 possible parent pairs. For sire parentage, the most likely sire was identified and the LOD score stored (mLOD1), along with the difference between mLOD1 and the LOD score for the next most likely sire. This difference was referred to as Δ1. The LOD score for the second most likely sire (mLOD2) and associated Δ2 were stored. The same method was used for dam parentage and for sire-dam parentage.

For Δ, a threshold (TΔ) was declared at TΔ = 3, and was used in all flocks. Parentage was only assigned if the most likely parent was at least 3 times more likely than the second most likely parent. Given the threshold TΔ = 3, a threshold for mLOD, (TmLOD) was found that balanced the number of false positives (i.e. mLOD2 > TmLOD) and false negatives (i.e. mLOD1 < TmLOD), subject to the constraint that the percentage of false positives was less than 10%.

For the real lambs, mLOD and Δ were compared to the thresholds TmLOD and TΔ, and parentage assigned if mLOD ≥ TmLOD and Δ ≥ TΔ, or not assigned if mLOD < TmLOD and Δ < TΔ. In all simulations and analyses we assumed a genotyping error rate of 1%.

RESULTS

Table 3. Assignment rates (AR %) for real data, False negative (-ve %), False positive (+ve %) and TmLOD (simulated data) using varying number of SNP and panels - 127 (W12), 190 (W123) or 191 (W126)

Panel	W12				W123				W126			
	AR	Tm LOD	-ve	+ve	AR	Tm LOD	-ve	+ve	AR	Tm LOD	-ve	+ve
Flock 1 Sire	97.3	5.2	1.7	1.7	98.3	9.4	0.7	0.8	99.5	7.1	0.0	0.4
Flock 1 Dam	88.2	7.7	3.7	1.6	94.7	10.7	0.7	0.7	95.4	10.2	0.8	0.9
Flock 1 Trio	96.4	22.8	0.8	1.2	97.8	33.7	0.1	0.2	98.6	33.0	0.2	0.5
Flock 2 Dam	81.4	5.9	5.4	5.0	81.4	9.7	0.9	2.1	90.7	9.1	2.5	2.2
Flock 3 Sire	80.6	4.6	0.9	0.8	86.4	4.7	0.0	0.0	79.6	9.8	0.0	0.2
Flock 3 Dam	81.6	6.1	1.2	0.8	91.3	6.7	0.1	0.1	88.3	8.0	0.1	0.2
Flock 3 Trio	77.7	19.6	0.1	0.5	78.6	33.3	0.0	0.0	78.6	30.1	0.0	0.1
Flock 4 Sire	49.4	2.1	1.1	1.2	48.3	4.3	0.8	0.5	48.3	30.3	0.5	0.2
Flock 4 Dam	60.0	4.3	1.7	1.3	62.2	4.6	0.2	0.2	60.0	7.8	0.4	0.9
Flock 4 Trio	28.3	18.4	0.5	0.4	28.9	32.8	0.2	0.2	28.9	30.3	0.1	0.1

False negative and false positive rates decreased (false positives decreased from 1.7% to 0.8% in Flock 1 sires simulated data) when changing from W12 to W123. This trend was evident across most flocks and groups (sire, dam or trio). In the real data, generally the assignment rate increased as the number of panels changed from 2 to 3. The exception is Flock 4, for which assignment rates were lower in some of the groups. There was not a significant difference in assignment rate across the panels examined (p=0.95). Upon investigation, it was discovered that 40% of dams and 50% of sires of Flock 4 had not been genotyped for this particular study. The number of genotypes for dams in this flock was the smallest across all the flocks at 21, but the use of embryo transfer in this flock has allowed the dams to have a large number of progeny (range of 0-15, average of 5).

DISCUSSION

This study demonstrated that a small number of SNP panels (2-3) generate adequate parentage assignment rates in Australian sheep flocks. The results also indicate that the design of the SNP panels is technically robust. Their performance across 4 flocks and 2 breeds in this study showed high rates of assignment where genotypes of potential sires, dams and lambs were available. Assignment rates were lower when fewer sires and dams within that flock were genotyped. The parentage assignment methodology developed allows for the assumption of a genotyping error rate, and it can be set to account for the observed error rate in any given SNP genotyping platform. This prevents the true sire from being eliminated on the basis of a single genotyping error. Importantly, the approach also uses the allele frequency at each SNP within the flock to generate population specific thresholds. This ensures that the thresholds for assignment are specific for each breed and flock.

The availability of parentage SNP panels with inbuilt performance SNPs is attractive to industry. The benefits for producers can include obtaining additional information such as the Poll status of the animal at no additional cost. The candidate SNPs that comprise the performance SNPs in parentage panels will be an area of focus for future research.

Utilising 2 panels of multiplexed SNP (or 127 loci) gave high rates of correct parentage and may be sufficient for many flocks, but 3 panels (or 190 SNP) provided higher confidence and is the recommendation for initial commercial application of a DNA based parentage testing product for less than AUD20.

ACKNOWLEDGEMENTS

The authors wish to thanks Felice Driver and MLA for financial support. Darryl Irwin (Sequenom, Queensland) performed multiplex assay design and testing. Julius van der Werf contributed to the methods development and assisted in the transfer of the working SNP panels to Daniel Pomp (GeneSeek) for production genotyping.

REFERENCES

Kalinowski S.T., Taper M.L. and Marshall T.C. (2007) *Molecular Ecology* **16**(5):1099.
Kalinowski S.T., Taper M.L. and Marshall T.C. (2010) *Molecular Ecology* **19**(7):1512.
Kijas J.W., Townley D., Dalrymple B.P., Heaton M.P., Maddox J.F., McGrath A., Wilson P., Ingersoll R.G., McCulloch R., McWilliam S., Tang D., McEwan J., Cockett N, Oddy V. H., Nicholas F.W. and Raadsma H. for the International Sheep Genomics Consortium. (2009) *PLoS ONE* **4:** e4668.
Kijas J.W., Lenstra J., Hayes B., Boitard S., Porto Neto L., San Cristobal M., Servin B., McCulloch R., Whan V., Gietzen K., Paiva S., Barendse W., Ciani E., Raadsma H., McEwan J., Dalrymple B., and the International Sheep Genomics Consortium (2012) *PLoS Biology* **10**(2):e1001258.
Marshall T.C., Slate J., Kruuk L.E. and Pemberton J.M. (1998) *Molecular Ecology* **7**(5):639.