SNiP and cut: Quantifying the potential benefits of genomic selection tools for genetic fault elimination in sheep

J.A. SISE[1], B. AUVRAY[2], K.G. DODDS[2] and P.R. AMER[1]

[1]AbacusBio Ltd., P.O. Box 5585, Dunedin, New Zealand
[2]AgResearch Invermay, Private Bag 50-034, Mosgiel, New Zealand

ABSTRACT

Whole genome single nucleotide polymorphism marker (SNP) technologies are rapidly being developed and could well play a role in the future selection of farmed animals. In sheep, work is being carried out to identify SNP's associated with economically important production traits and disease resistance. Stud sheep breeders routinely cull up to 30% of lambs born due to a variety of faults. Despite this culling pressure, the faults keep occurring, suggesting low heritability, recessive inheritance or incomplete penetration of fault causing genes. The objective of this paper was to identify the potential of SNP fault detection and elimination in sheep. Simulation was used to predict changes in gene frequencies and in an index of genetic merit of production traits over time when between two and 30 SNPs were used to aid selection. The SNPs are assumed to predict the presence of deleterious recessive genes. As selection pressure applied to individual or combination SNPs increased, the rate of increase in production trait genetic merit slowed down. Thus, a balance would be required between the emphasis on SNPs actively used to select against genetic faults, relative to emphasis on genetic merit. This work identified scope for substantial economic benefits from application of SNP technology for removal of faults to both stud breeders and commercial sheep farmers.

Key words: single nucleotide polymorphism; sheep; breeding.

INTRODUCTION

Culling of lambs due to conformational faults is routine practice in most sheep farms. A typical New Zealand stud sheep flock can cull up to 30% of lambs which are judged unsuitable for breeding (S.F. Glennie, Personal communication). Traditionally, ram breeders have seen merit in eliminating animals with problems such as undershot jaws, bad feet or black spot (Fleet *et al.*, 2002), but with increased selection for production traits (Amer, 2000) many of these faults are still present in the breeding populations. Additionally, where these faults are caused by recessive genes or additive genes displaying incomplete penetration, then identification and elimination of carriers becomes difficult using traditional selection methods. An example of a recessive fault causing gene is the microphthalmia gene (Jolly *et.al.*, 2004). This gene, estimated to be present in 5 to 10% of the Texel and Texel-cross rams in New Zealand causes blindness in lambs. A commercial DNA test is available to detect microphthalmia carriers, prior to breeding (Dodds, 2007). If DNA technologies allowed identification of likely carriers, fault occurrence could be significantly reduced or removed from breeding and commercial populations.

Recent advances in molecular genetics have enabled the development of technologies for genotyping with multiple DNA markers simultaneously. In humans, genetic research has a substantial focus on fault identification, with genome-wide association studies used to identify at risk loci for diseases such as Taybi_Linder syndrome (Leutenegger *et al.*, 2006). New techniques developed for fault identification in humans are likely to be applicable to animal research. In New Zealand, AgResearch has a large research programme, funded by Ovita Ltd, which is aimed at identifying sheep single nucleotide polymorphism markers (SNPs). If ovine SNP tests could be identified for fault genes, lambs carrying the undesirable SNP allele could be penalised at selection to reduce the frequency of the fault causing gene. Potentially this could improve the efficiency of both stud and commercial sheep breeders, by increasing the number of lambs suitable for breeding or as commercial ewe flock replacements.

The objective of this paper is to present a SNP system simulation model of a sheep breeding programme, and use it to quantify the benefits of potential selection programmes focusing on genetic merit and the elimination of faults.

MATERIALS AND METHODS

A simulation model has been developed to predict changes in allele frequencies and the effect on genetic progress in sheep flocks through selection against SNPs linked to faults. The model has been developed using Mathcad 14.0 software.

The simulation of changes in bi-alleic SNP frequencies over time was based on the methodology of Kearney *et al.* (2005). Let a_l be the proportion of breeding rams of age l, and $b_{j,l,t}$ be the

proportion of rams of genotype *j*, age *l* at time *t*. The proportion of rams of genotype *j* mated at time ($s_{j,t}$) can be calculated as:

$$s_{j,t} = \sum_{\text{min ram-age}}^{\text{max ram-age}} a_l b_{j,l,t}$$

An identical calculation can be used for determining the proportion of ewes of genotype k at mating:

$$d_{k,t} = \sum_{\text{min ewe-age}}^{\text{max ewe-age}} c_l d_{k,l,t}$$

where c_l is the proportion of breeding ewes of age *l*, and $d_{k,l,t}$ is the proportion of ewes of genotype *k*, age *l* at time *t*, and 2 and 6 are the minimum and maximum ages of ewes mated respectively.

The proportion of offspring of genotype *i* at time of birth ($o_{i,t}$) assuming random mating can be calculated as:

$$o_{i,t} = \sum_{j=1}^{n} \sum_{k=1}^{n} P(o_{i,t} \mid S_j, D_k) s_{j,t} d_{k,t}$$

where *n* is the number of genotypes for the gene of interest, and $P(o_{i,t}|S_j,D_k)$ is the probability of an offspring of genotype *i* at time *t*, given a sire of genotype S_j and a dam of genotype D_k, where the $s_{j,t}$ and $d_{k,t}$ represent the proportion of sires of genotype *j* and dam genotype *k*, in the year of mating (t).

The simulation model generates results for two to 30 SNPs for 30,000 of both ram and ewe lambs. For the *zth* SNP, the offspring genotype SNP_z (where SNP_z = 0, 1 or 2 as the number of copies of the SNP favourable allele) is simulated by mating animals randomly drawn from a selected parent population. Genotype frequencies within the selected parent population are specified in a separate computational step. This step takes SNP results for lambs born in a given year, and applies selection to them using an aggregate score containing multiple SNP information combined with the simulated index of production trait breeding values. Selected individuals enter the parent population each year, and the average genotype frequencies for the new parent population are used in the equation for $o_{i,t}$.

An index of estimated breeding values defining aggregate production trait merit was simulated for each lamb by sampling index values from a random normal distribution with a mean equal to the index average of selected parents mated, and a standard deviation equal to the square root of the assumed heritability (0.3) of the index. The mean index value of sires and ewes in the initial population was set to zero. Subsequently, index values of selected lambs that become parents were calculated assuming that each year the top 5% of males and 70% of females are selected to enter the parent pool, on an index of production trait breeding values and an aggregate of SNP scores.

Simulated ram and ewe lambs in stud breeding flocks are selected according to a combined SNP and index result *r*. For each SNP *z* generated, the simulated lamb is given a SNP_z score. A weighting factor β_z is used to define how important SNP_z is relative to other SNPs under selection. As β_z increases (range = 0.06 to 0.5), more selection pressure is applied to SNP_z, and animals carrying favourable alleles are more likely to be selected than animals carrying unfavourable alelles. Results for each lamb r_n can be calculated:

$$r_n = i_n + \sum_{z=1}^{p} \beta_z SNP_z$$

where i_n is the lamb simulated estimated breeding value for the production trait index, and p is the number of SNP's used in the selection.

Phenotypic selection was also modelled using selection against homozygote carriers of the unfavourable allele. A SNP_z score of -2 was applied to homozygous carriers with no preference given to carriers of the favourable allele. This effectively means that all carriers of two copies of unfavourable

TABLE 1: Changes in the frequency of the unfavourable allele in lambs due to variable initial allele frequencies (set A), variable amounts of selection pressure (set B) and phenotypic selection only (set C).

Set		Year			
		1	2	3	4
A	SNP1	0.40	0.18	0.05	0.01
A	SNP2	0.30	0.13	0.04	0.01
A	SNP3	0.20	0.08	0.02	0.00
A	SNP4	0.10	0.04	0.01	0.00
B	$\beta = 0.5$	0.40	0.20	0.07	0.02
B	$\beta = 0.125$	0.40	0.29	0.19	0.11
B	$\beta = 0.06$	0.40	0.34	0.27	0.21
C	Phenotypic selection	0.40	0.31	0.22	0.17

TABLE 2: Affect of β on lamb index value and genetic progress relative to no selection, when four SNPs are simultaneously selected with the initial frequency of unfavourable alleles set to 0.4.

Selection method	Index value (by year)			
	1	4	8	12
$\beta = 0.5$	0.084	0.420	0.976	1.637
$\beta = 0.125$	0.165	0.716	1.389	2.088
$\beta = 0.06$	0.181	0.771	1.485	2.203
Phenotypic selection	0.158	0.661	1.279	1.952
$\beta = 0$	0.184	0.792	1.523	2.264
	Relative genetic progress (%)			
$\beta = 0.5$	46	53	64	72
$\beta = 0.125$	90	90	91	92
$\beta = 0.06$	98	97	98	97
Phenotypic selection	86	83	84	86

FIGURE 1: Allele frequency of the unfavourable allele in rams and ewes mated, and lambs born over 12 years of selection, using five SNPs all with initial unfavourable allele frequencies of 0.4 and selected using β=0.5. ● = Ram lambs selected; ○ = Rams mated; ■ = Ewe lambs selected; □ = Ewes mated.
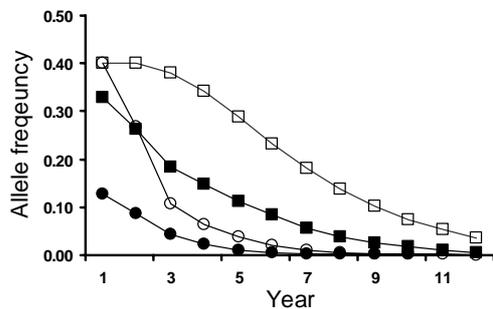
SNP's are identified by the breeder and culled.

The population structure used to simulate SNP and index values for parents mated assumed generation intervals of 1.5 and 3.5 years for rams and ewes respectively. Initial SNP genotype allele frequencies are assumed to be constant over age groups for rams and ewes mated. SNP allele frequencies in the initial population were pre-set to defined allele frequencies, where the frequency of the unfavourable allele ranged from 0.05 to 0.4, or simulated according to a random uniform distribution where the frequency of the unfavourable allele was assumed to be between 0.3 and 0.4.

RESULTS

Effect of SNP selection on the frequency of unfavourable alleles in stud breeding programmes

Figure 1 shows the simulation results for changes in the unfavourable allele frequencies for rams and ewes mated, lambs born, and ram and ewe lambs selected for breeding over 12 years, when 5 independent SNPs were simulated. When fault linked SNP markers are used to assist breeding programme selection decisions, high selection pressure on rams results in a rapid decrease in the allele frequency of the unfavourable allele. With the initial frequency of the unfavourable allele of each SNP in the parent population set to 0.4 and β set to 0.5 for a high SNP selection emphasis, identical results are observed for each of the five SNPs. With just 5% of ram lambs required for breeding, the allele frequency of the unfavourable allele in ram lambs selected in the first year is 0.13. Ewe lambs are selected where an average frequency of the unfavourable allele in the top 70% is 0.33. Within seven years, the frequency of the unfavourable allele in lambs born has dropped from 0.4 to less than 0.1 resulting in just 1% of lambs born with two copies of each or any unfavourable allele.

The rate of decrease in the frequency of unfavourable alleles is dependent on both the initial allele frequency and the selection pressure applied (β). Table 1 shows the changing frequency of the unfavourable allele in lambs for three different scenarios over 12 years of selection. In data set A, four SNPs with decreasing initial allele frequencies are simultaneously selected using β = 0.5. In data set B, four SNPs with the same initial allele frequencies are simultaneously selected using decreasing levels of β. In data set C, homozygous carriers of four SNPs with the same initial allele frequencies are selected against using β = 0.5, representing selection for phenotype, not genotype. In dataset A, as the initial frequency of the unfavourable allele drops in the lamb population, the rate of decrease also declines, resulting in more effective pressure on the high frequency SNPs. In dataset B, as selection pressure decreases, the rate of decline also decreases. The high initial frequencies of all SNPs used resulted in a slower decrease in frequency of unfavourable alleles than for SNP1 in set A, even for the SNP with β = 0.5. In data set C, removal of homozygote carriers using phenotypic selection, results in a decrease in the frequency of the unfavourable allele for all four SNPs. However, because heterozygous carriers are not penalised at selection, the frequency of the unfavourable allele remains high in comparison to selection by genotype.

Effect of SNP selection on production trait genetic progress in stud breeding programmes

Table 2 shows the effect of β on simulated index values of lambs born, where 4 SNP's have been used at varying amounts of SNP selection pressure. As SNP selection pressure goes up, the rate of increase in production trait index values decreases. If selection pressure on the SNPs is high, then animals with high index values may be penalized for having one or more unfavourable SNP alleles and may not be selected as flock replacements. The relative genetic progress made by SNP and phenotypic selection methods compared to selection on index value alone is also shown. Removal of homozygote fault carriers by phenotypic selection results in around 85% of the genetic progress that would be made on production traits if selection was on index value alone. Less genetic progress is sacrificed when low SNP selection pressures (β = 0.125, β = 0.06) are used.

Effect of SNP panel in reducing the incidence of animals exhibiting homozygous recessive faults

In order to assess the impact of detection of low frequency SNPs in a breeding population, 15 SNPs were simulated where the initial frequency of the unfavorable allele was randomly uniformly distributed between 0.3 and 0.4. Using a constant selection pressure of β = 0.25, the average frequency

of the unfavourable allele dropped from 0.35 to 0.21 in just five years. For a recessive inherited trait, this would more than halve the number of lambs born that express the disorder. After 13 years of selection the average frequency of the unfavourable alleles was less than 5% for all SNP markers.

DISCUSSION

We have extended the model of Kearney *et al.* (2005) to accommodate simultaneous simulation on large numbers of individual SNPs in conjunction with an index of production trait values. Other sheep simulation models incorporating marker and production trait information (e.g. Amer *et al.*, 2005) are much slower to run, and in practice are limited to a relatively small number of specific genetic loci. The model is easily scalable to simulate large numbers of SNPs, with a modest requirement for computing time and resource.

In order to assess the financial impact of fault detection using SNP technologies we have estimated the impact on a standard dual purpose selection index (Amer, 2000) of reducing the proportion of lambs culled from 30% to 15%. In a typical 1,000 ewe flock weaning 140% lambs there are 700 ram and 700 ewe lambs born. If 30% of these lambs are culled due to faults then only 490 ewe and 490 ram lambs are available for selection. If the breeder needs to replace 25% of ewes each year, they require 51% of the available ewe lambs. Reducing the cull rate from 30% to 15% results in the available replacements increasing from 490 to 595, allowing selection of the top 42% of ewe lambs on merit. On the ram side, the breeder wishes to choose 25 ram lambs. With 490 ram lambs available after 30% culling, the top 5% of ram lambs are chosen. If 595 ram lambs are available after culling, the same breeder can choose the top 4% of rams.

A reduction in the number of culls from 30% to 15% would increase the selection index value of an average lamb from 79 cents per year to 86 cents per year. This translates to an additional 8% of genetic progress, but does not account for genetic progress lost in production traits. At low to moderate levels of selection emphasis on unfavourable SNPs, loss in genetic progress in the production traits ranges from 2 to 10%, but culling animals with observed faults is also costly and less effective in the medium to long term. In addition, because less ram lambs are culled, there are more lambs available to sell to ram buyers. If a breeder is selling from the top 50% of their rams, with 700 ram lambs born, there are 270 and 220 ram lambs of the same genetic production merit at 15% and 30% culling respectively.

Additional benefits can also be accrued by commercial farmers, through purchase of rams with less faults. This would allow commercial farmers to maximize the number of ewes mated to terminal sires, with more ewe lambs born to maternal sires suitable for use as flock ewe replacements. In a commercial ewe flock with 1,000 ewes, a 25% replacement rate and 30% of lambs culled, the farmer needs to mate 51% of the ewes to maternal sires to ensure adequate flock ewe replacements, compared to 45% if 20% of ewe lambs were culled. This equates to an additional 84 terminal lambs weaned. Assuming the carcase premium for a terminally sired lamb is $3 higher than a maternally sired lamb, the additional 84 terminal lambs result in $252 added value.

With discernible benefits for both ram breeders and commercial farmers, the challenge for SNP assisted fault detection lies in development of a cost effective test. If the fault associated SNPs are linked to recessive or additive genes with incomplete inheritance, then the opportunity to identify and validate the SNPs within multiple sheep breeds may be limited with current discovery strategies. However, given the opportunity, selection using SNPs to eliminate faults could provide significant financial gain to sheep breeders and commercial sheep farmers.

ACKNOWLEDGEMENTS

REFERENCES

Amer, P.R. 2000: Trait economic weights for genetic improvement with SIL. *Proceedings of the New Zealand Society of Animal Production* **60**: 189-191.

Amer, P.R.; Sherlock, R.G.; Tate, M.L. 2005: Using assortative matings to maximize sales of homozygous sheep sires, a simulation study. *Proceedings of the Association for the Advancement of Animal Breeding and. Genetics* **16:** 64-67.

Dodds, K.G. 2007: Use of gene markers in the New Zealand sheep industry. *Proceedings of the Association for the Advancement of Animal Breeding and Genetics* **17**: 418-425.

Fleet, M.R.; Mahar, T.J.; Turk J.A. 2002: Merino crossbreeding and objectionable sheep fibres: the problem and potential solution. *Wool technology and sheep breeding* **50**: 650-656.

Jolly, R.; Blair, H.T.; Johnstone, A.C. 2004*:* Genetic disorders of sheep in New Zealand: A review and perspective. *New Zealand veterinary journal* **52(2)**: 52–64.

Kearney, J.F.; Amer, P.R.; Villanueva, B. 2005: Cumulative discounted expressions of sire genotypes for the complex vertebral malformation and b-casein loci in commercial dairy herds. *Journal of dairy science* **88**: 4426-4433.

Leutenegger, A.L.; Labalme, A.; Genin, E.; Toutain, A.; Steichen, E.; Clergot-Darpoux, F.; Edery, P. 2006: Using genomic inbreeding coefficient estimates for homozygosity mapping of rare recessive traits: Application to Taybi-Linder syndrome. *The American journal of human genetics* **79**: 62-66.

McEwan, J.C. 2007: Developing genomic resources for whole genome selection. *Proceedings of the New Zealand Society of Animal Production* **67***:* 148-153.