Proc. Assoc. Advmt. Anim. Breed. Genet. 20:179-182

COMPARISON OF THE POWER OF POOLED GENOTYPING STRATEGIES TO DETECT SIGNIFICANT SNP EFFECTS FOR FLYSTRIKE RESISTANCE IN MERINO SHEEP

S. Dominik, J.M. Henshall and J. Smith

CSIRO Animal, Food and Health Sciences, New England Highway, Armidale, NSW 2350

SUMMARY

Genotyping pooled DNA is a cost-effective strategy to produce genotype information for whole genome association studies. The objective of this study was to compare the power to detect significant SNP effects for flystrike of simulated genotyping strategies in the Breechstrike Resource flock. A gene dropping approach was used to simulate allele frequencies. Individual genotyping was used to set a benchmark for comparison with three DNA pooling strategies. These included pooling before fixed effect adjustment, pooling after fixed effect adjustment and case-control pooling. The study showed that the highest power to detect significant associations between SNP allele frequencies and phenotypes can be achieved by individual genotyping. Case-control pooling and pooling after fixed effect adjustment had similar power to detect significant SNP effects, whereas pooling before fixed effect adjustment performed worst. The high power of detection of SNP effects of the individual genotyping strategy indicates that the Breechstrike Resource flock is a suitable resource for the detection of significant SNP effects for flystrike. However, pooling affects the power of detection of significant SNP effects, in particular when effects are small.

INTRODUCTION

Fly strike in Merino sheep, where fly larvae feed into the tissue, is a welfare issue in the Merino sheep industry. Traditionally, mulesing has been used as a strategy to prevent fly strike. Important and difficult to measure phenotypes, such as fly strike resistance, are ideal targets for genomic approaches. Whole genome association studies are most powerful where very large numbers of individuals are genotyped, however, individual genotyping is very costly. There is growing evidence that pooled DNA can be used successfully. This was mostly demonstrated for binomial phenotypes (Lee 2005, Huang *et al*. 2010), however, it has also been shown that DNA pooling is an effective strategy for quantitative traits (Henshall *et al.* 2012). Whilst pooled genotype strategies are more cost-effective than individual genotype strategies, the trade-off is a loss of power to detect SNP effects. The objective of this study was to compare simulated DNA pooling strategies with individual genotyping for their power to detect significant SNP effects for flystrike in the Breechstrike Resource flock.

MATERIALS AND METHODS

Data. Phenotype and pedigree data of the Breechstrike Resource flock (Smith 2009) were used for this study. The pedigree contained 3109 individuals, including 463 base animals born in 2003 and 2646 progeny born between 2005 and 2011, of which 2274 have flystrike phenotypes. Founder animals originated from three genetic groups: Ultrafine/Superfine (US), Fine/Fine Medium (FFM) and Medium/Strong (MS). The genetic groups provided the structure for the gene dropping approach used in this study. Contemporary groups were formed by sex (male or female), birth year (2005 – 2011) and mulesing status (mulesed or unmulesed), which were recorded on each animal that had a flystrike record. Phenotypes included flystrike (struck / not struck), wrinkle (high (H), moderate (M) and low (L)) and wool cover on the breech (high (H), moderate (M) and

low (L) breech cover). Flystrike, wrinkle and breech cover phenotypes formed phenotype classes. The combination of phenotype and contemporary groups was used for the DNA pool assignment.

Gene dropping. Allele frequencies for each genetic founder group (US, FFM and MS) were simulated using gene dropping (MacCluer et al. 1986). Single nucleotide polymorphism (SNP) genotypes were assigned at random to base animals, and then transmitted through the pedigree subject to Mendelian inheritance rules, i.e. progeny have equal chance of inheriting each of the two alleles carried by the parent. One hundred SNP were simulated and the gene dropping procedure was repeated 100 times. Simulations were run with and without an existing association between alleles and phenotypes. When an association between SNP alleles and phenotype was simulated, SNP effect sizes ranged from 0.1 to 1 phenotypic standard deviation (σ_p).

Genotype pooling strategies. Three pooling strategies were tested and compared to individual genotyping (Strategy 1). Individual genotyping of 2274 animals served as a benchmark for the other three strategies. In Strategy 2, animals were pooled for genotyping within phenotype (flystrike – struck/not struck, wrinkle, breechcover) and contemporary group (birthyear, sex and mulesing status). The numbers and size of pools resulting from Strategy 2 are shown in Table 1. In Strategy 3, flystrike phenotypes were adjusted for fixed effects (birthyear, sex and mulesing status). Individuals for genotyping were pooled with the objective of creating a balanced number of pools in the struck and not struck group and achieving even pool sizes (number of individuals per pool) across phenotype / contemporary groups (Table 1). Strategy 4 is a case-control pooling approach and uses a combination of individual and pooled DNA for genotyping. All struck animals are individually genotyped and matched, if possible, with an individual genotype of a not struck animal from the same contemporary / phenotype group. Not struck animals that were not paired with a struck animal are pooled within contemporary / phenotype group. Numbers and sizes of pools resulting from Strategy 4 are summarised in Table 1.

Table 1. Numbers of pools and pool size resulting from three pooling strategies; Strategy 1 being individual genotyping.

	Pool size	Number of pools		
	Number of individuals	Total	Struck	Not struck
Strategy 2				
	1	318	183	135
	5	174	20	154
	6	181	15	166
Strategy 3				
	1	373	373	0
	5	379	0	379
	6	1	0	1
Strategy 4				
	1	717	373	344
	$\leq$ 10	34	0	34
	11-20	21	0	21
	> 20	29	0	29

Analysis. Associations between flystrike phenotypes, “struck” and “not struck”, and allele frequencies were established by logistic regression. The phenotype and contemporary groups were included in the model as fixed effects. Analysis were conducted with software written in R (R Development Core Team 2008)

RESULTS AND DISCUSSION

For simulations without associations between phenotypes and allele frequencies, no more significant SNP than expected by chance were found with any of the genotyping strategies at significance levels P < 0.05, P < 0.01 and P < 0.001 (Table 2).

Table 2. Relative frequency of detection (%) of SNPs associated with the phenotype at three significance levels; no association between phenotype and allele frequency was simulated

Significance level	Relative frequency of detection (in %)			
	Strategy 1	Strategy 2	Strategy 3	Strategy 4
< 5%	3.60	3.30	3.60	3.40
< 1%	1.10	0.30	0.40	0.40
< 0.1%	0.03	0.01	0.02	0.01

When SNP effects of varying size were simulated to be associated with allele frequencies, associations were detected more often than just by chance (Figure 1). For all strategies the power to detect SNP effects increased with increasing SNP effect size. In reality most SNP effects are small and pooling strategies 2, 3 and 4 were not very powerful in detecting them. Strategy 1 (individual genotypes) had the highest power to detect small SNP effects ($0.1\sigma_p$), with a relative frequency of 18% (P < 0.05), whereas with Strategy 2, 3 and 4 ranged from 5.5% to 7%. Strategies 3 and 4 yielded similar results, with Strategy 3 being slightly more powerful (Figure 1). Strategy 2 was the least powerful approach at detecting significant SNP associations with phenotypes compared to all other genotype pooling strategies.

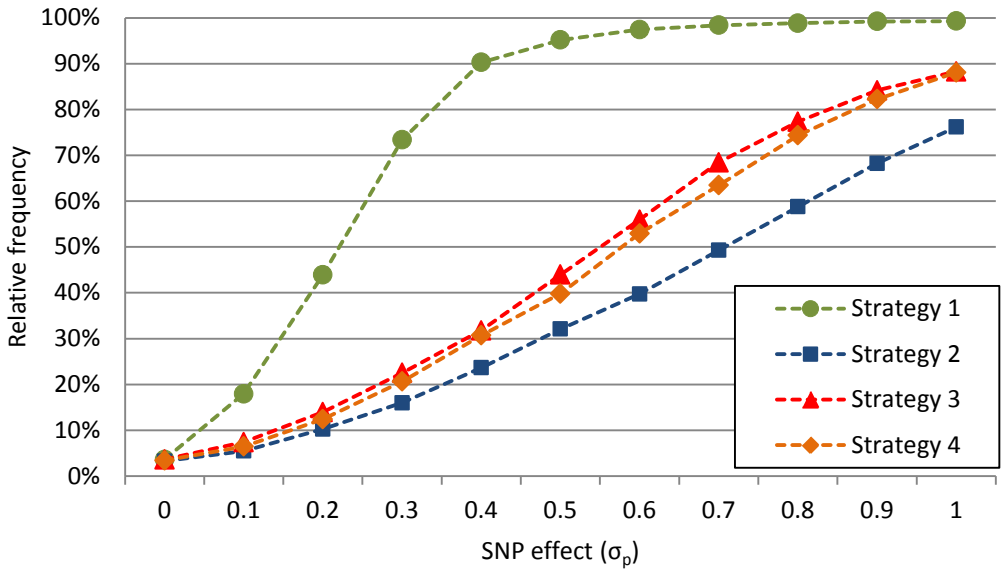

Figure 1. Relative frequency of detection of significant SNP effects (in %) of varying size (x axis in σ_p) with four different genotyping strategies at significance level P < 0.05.

Figure 2 shows the frequency of detection of SNP effects for Strategies 1 and 3 to demonstrate the difference of detection at different significance levels. The pooled genotyping strategies use

between 65-70% fewer assays compared to individual genotyping and the power of detecting significant SNP effects of small size reduced by 64-78% with small effects up to $0.4\sigma_p$ and 11-31% with large SNP effect sizes of $0.8\text{-}1\sigma_p$. Power was lower than expected based on results presented by Henshall *et al.* (2012). Huang *et al.* (2010) suggested increasing pool sizes as much as possible to estimate allele frequencies accurately, but this was not possible in this study due to small number of individuals in each phenotype / contemporary group class.

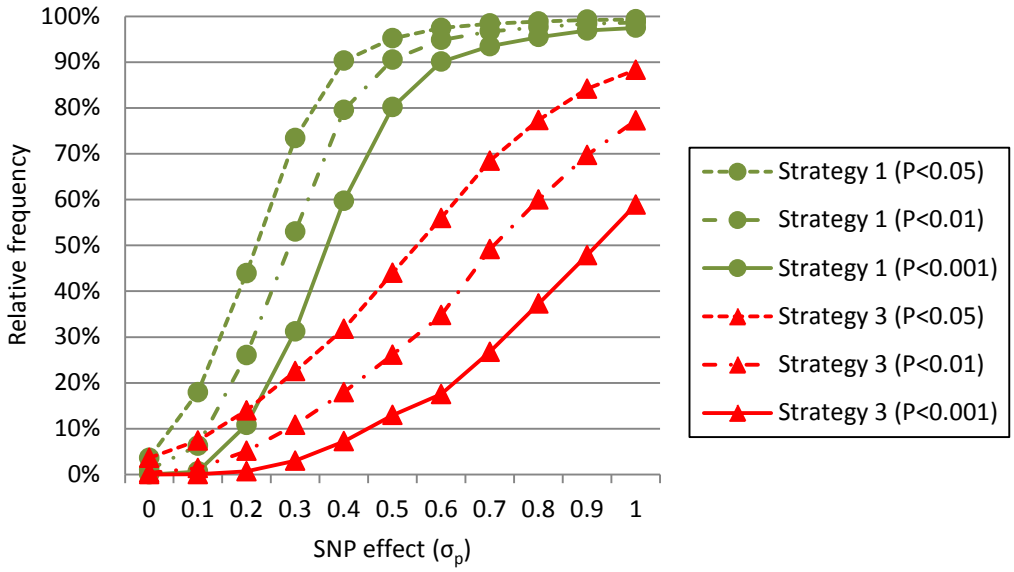

Figure 2. Relative frequency of detecting significant SNP effects of varying size (x axis in σ_p) with two pooling strategies 1 and 3 at three significance levels.

CONCLUSIONS

The power of DNA pooling approaches is affected by the nature of the phenotype and the number of contemporary groups in the data set. Pooling strategies lack power in the detection of small SNP effects; however, they could still provide a cost-effective alternative for the estimation of genomic breeding values. Pooling strategies should be designed and tested for their power prior to implementation.

ACKNOWLEDGEMENTS

This work was funded by Australian Wool Innovation Ltd.

REFERENCES

Henshall J.M., Hawken R.J., Dominik S. and Barendse W. (2012) *Genet. Sel. Evol*. 44: 12.
Huang W, Kirkpatrick BW, Rosa GJM, Khatib H. (2010) *Anim. Genet*. **41**: 570.
Lee W.C. (2005) *Cancer Epidem. Biomar*. **14**: 958.
MacCluer J.W., Vandeburg J.L., Read B. and Ryder O.A. (1986). *Zoo Biol.* **5**:149.
R Development Core Team (2008) 'R: A language and environment for statistical computing.' R Foundation for Statistical Computing, Vienna, Austria.
Smith J.L., Brewer H.G. and Dyall T. (2009) Proc. Assoc. Advmt. Anim. Breed. Genet. **18**:334.