Proc. Assoc. Advmt. Anim. Breed. Genet. 20:412-415

PARTITIONING THE GENETIC VARIANCE INTO GENOMIC AND PEDIGREE COMPONENTS FOR PARASITE RESISTANCE IN SHEEP

M. Al Kalaldeh J.P. Gibson, J.H.J. van der Werf and C. Gondro

University of New England, School of Environmental and Rural Science,
Armidale 2351, Australia

SUMMARY

In this study, we estimated the additive genetic variance explained by genomic markers for parasite resistance in a large mixed population of sheep and compared this estimate to the additive genetic variance explained by pedigree. Furthermore, we partitioned the total genetic variance by fitting both of genomic relationship matrix (GRM) and numerator relationship matrix (NRM) simultaneously into a genomic component explained by genomic relationships and a polygenic component explained by pedigree relationships. In this analysis, all the genetic variation explained by pedigree could be captured by the 50K SNP chip markers. When both of GRM and NRM were fitted simultaneously, 73.7% of total genetic variance was explained by genomic effects while the remaining variance (26.3%) was explained by pedigree effects. The proportion of genetic variance explained by genomic effects was further partitioned into 26 chromosomes. A significant relationship was found between chromosome-specific variance and the length of the chromosome ($R^2 = 0.26$). This indicates that disease resistance is a largely polygenic trait with a large number of genes involved in the mechanisms of resistance but there are some chromosomal regions that explain a larger proportion of the variation.

INTRODUCTION

Parasite resistance for nematode infection is a complex trait of great importance in sheep and other livestock species. Breeding for sheep resistance is a viable method to reduce the effect of these nematodes on production and to reduce the cost of anthelminitic treatments (Dominik 2005). The identification of genes or genomic regions associated with sheep resistance would greatly accelerate genetic improvement in breeding programs. To date, genome wide association studies (GWAS) for parasite resistance have identified genetic variants that together explain only a small proportion of genetic variance of the trait (Kemper *et al.* 2012). Recently, Yang *et al.* (2010) showed that a considerable proportion of genetic variance can be explained by considering all single-nucleotide polymorphism SNPs simultaneously in a mixed linear model analysis. This mixed model has the potential to accumulate the effects of associated SNPs that might be too small to pass the significance threshold of single-SNP GWAS analysis.

To investigate in more details the role of SNP markers in parasite resistance, we used data from a large mixed breed population of sheep naturally challenged with *Haemonchus contortus*, and genotyped with the Illumina OvineSNP50 BeadChip. We estimated the additive genetic variance explained by genome-wide SNP data and compared this estimate to the additive genetic variance explained by pedigree. Furthermore, we partitioned the total genetic variance explained into genomic and polygenic components by fitting both of genomic data and pedigree simultaneously, and quantified the amount of genomic variance that can be explained by each chromosome.

MATERIALS AND METHODS

Animals and Phenotypes. Parasite resistance trait, as measured by WEC, was investigated in a multi-breed sheep population from the Sheep Cooperative Research Centre information nucleus flock (INF). A total of 7153 animals with both genotype and phenotype data were included in this analysis. Sires were either from Merino, terminal or maternal breeds and the size of resulting half-

sib families ranged from 20 to 91 with a median of 33 progeny. The breed content of the sheep population is shown in Table 1. Various breeds were represented in the population but with a significant proportion of Merino sheep, and only this breed had a substantial proportion of purebred animals. The remaining breeds were mainly represented by their crosses with Merino.

Table 1. Proportions of different breeds in the population

Breed	BL	COR	DH	SD	CO	PD	TX	AF	PER	PS	ME
Proportion	11.9	0.74	0.02	0.48	10.7	1.7	2.48	3.16	0.04	0.88	67.9

Border Leicester: BL, Corriedale: COR, Dorset Horn: DH, SD: Southdown, Coopworth: CO, Poll Dorset: PD, Texel: TX, Australian Finnsheep: AF, Perendale: PER, Prime Samm: PS, Merino:ME

Genotypes. Sheep were genotyped using the Illumina OvineSNP50 BeadChip. The following quality control measures were applied to the SNP data: SNPs were removed if they had a minor allele frequency (MAF) < 0.1%, a genotyping call rate < 90%, were not in Hardy-Weinberg equilibrium, and had no mapping information. The final data comprised genotypes for 47306 SNPs on 7153 animals.

Data analysis. The data were analyzed using the following mixed linear models:

Model1: $y^* = a + e$

Model2: $y^* = g + e$

Model3: $y^* = g + a + e$

Model4: $y^* = \sum_1^{26} g_i + a + e$

where y^* is a vector of adjusted phenotypic records, a is a vector of random additive genetic effects and assumed to be normally distributed with $N(0, A\sigma_a^2)$, A is the numerator relationship matrix (NRM) calculated from the pedigree data and σ_a^2 is the additive genetic variance, g is a vector of additive genetic effects accounted by all SNPs and assumed to be normally distributed with $N(0, G\sigma_g^2)$, G is the genomic relationship matrix (GRM) and σ_g^2 is the variance explained by all SNPs, g_i is a vector of additive genetic effects accounted by SNPs on the i^{th} chromosome and assumed to be normally distributed with $N(0, G_i\sigma_{gi}^2)$, G_i is the GRM built based on SNPs of the i^{th} chromosome, σ_{gi}^2 is the variance explained by SNPs on the i^{th} chromosome, and e is a vector of random residuals. The variance components were estimated using GCTA software (Yang *et al.* 2011).

Phenotypic records were adjusted for systematic environmental effects using the following model: $y = 1\mu + Xb + ZQa + e$, where y is a vector of cube root transformed WEC records, μ is the mean, X and Z are design matrices of fixed and random effects respectively, Q is a matrix containing breed proportions for each animal calculated from the pedigree records, b is a vector of fixed effects, a is a vector of random breed effects assumed to be normally distributed $\sim N(0, \sigma_q^2 I)$, where σ_q^2 is the variance of breed effects. The following fixed effects were included in the model: age at WEC recording, sex, rearing type, and contemporary groups formed using INF flock, group of management and year of birth.

RESULTS AND DISCUSION

The proportion of additive genetic variance explained by SNP markers or pedigree relative to the total variance corresponds to heritability of WEC. The estimated variance components from models 1to 4 are shown in Table 2. In this analysis, all the additive genetic variance explained by pedigree could be captured by the Ovine 50K SNP chip markers. This clearly indicates that the Ovine 50K SNP chip markers can trace all polygenic relationships due to sharing of causative

variants in this large mixed breed population of sheep.

Table 2. Genetic and genomic variances for WEC estimated in models 1 to 4

Model 1		Model 2		Model 3			Model 4
$\hat{\sigma}_a^2$	h_a^2	$\hat{\sigma}_g^2$	h_g^2	$\hat{\sigma}_a^2$	$\hat{\sigma}_g^2$	h_{a+g}^2	$\sum_{i=1}^{26} \sigma_{gi}^2$
0.67	0.147	0.67	0.147	0.20	0.56	0.1645	0.55

In model3, Both of GRM and NRM were fitted simultaneously in order to separate effects of pedigree (polygenic) relationships from genomic (SNPs) relationships. The total genetic variance estimated when both effects were fitted simultaneously was higher than the situation where each of them was fitted alone. Moreover, the residual (unexplained) variance of the total phenotypic variance was reduced in model 3 compared to the residual variance in model 1 and model 2. This indicates that there is not complete overlap between polygenic and genomic effects. In this model, a large proportion of total genetic variance was explained by genomic relationships (73.7%) while the remaining variance was explained by pedigree effects (26.3%).

In model4, the genomic variance explained by genomic relationships was further partitioned into 26 chromosomes. A GRM was built for each individual chromosome then all GRMs were fitted simultaneously to estimate the amount of genomic variance that can be attributed to each chromosome. The sum of estimates due to individual chromosomes was slightly lower than the genomic variance explained by genomic relationships in model 3. This suggests a very weak covariance between genomic relationships on different chromosomes.

A significant relationship between chromosomal length and the genomic variance explained by each chromosome (Figure 1 over page) is consistent with the hypothesis that many alleles with small effects contribute much of the genetic variation of the trait. It is notable, however, that five chromosomes exhibited higher contributions to genetic variance than expected given their size. This demonstrates that some chromosomal regions have effects larger than expected on a purely infinitesimal model.

In conclusion, our results suggest that the Ovine SNP50 array can capture a large proportion of genetic variance for WEC trait in a large multi-breed population of sheep. The same proportion of genetic variance can be attributed to individual chromosomes with a significant relationship between chromosomal length and the genomic variance explained by each chromosome.

REFERENCES

Dominik S. (2005) *Genet. Sel. Evol* **37**: S83.

Kemper K. E., Emery D., Bishop S. C., Oddy H., Hayes B. J., Dominik S., Henshall J. M. and Goddard M. E. (2011) *Genet Res Camb.* **93**: 203.

Yang J., Benyamin B., McEvoy B. P., Gordon S., Henders A. K., Nyholt D. R., Madden P. A., Heath A. C., Martin N. G., Montgomery G. W., Goddard M. E. and Visscher P. M. (2010) *Nature Genetics*. **42**: 565.

Yang J., Lee H., Goddard, M. E. and Visscher P. M. (2011) Am *J Hum Genet*. **88**: 76.

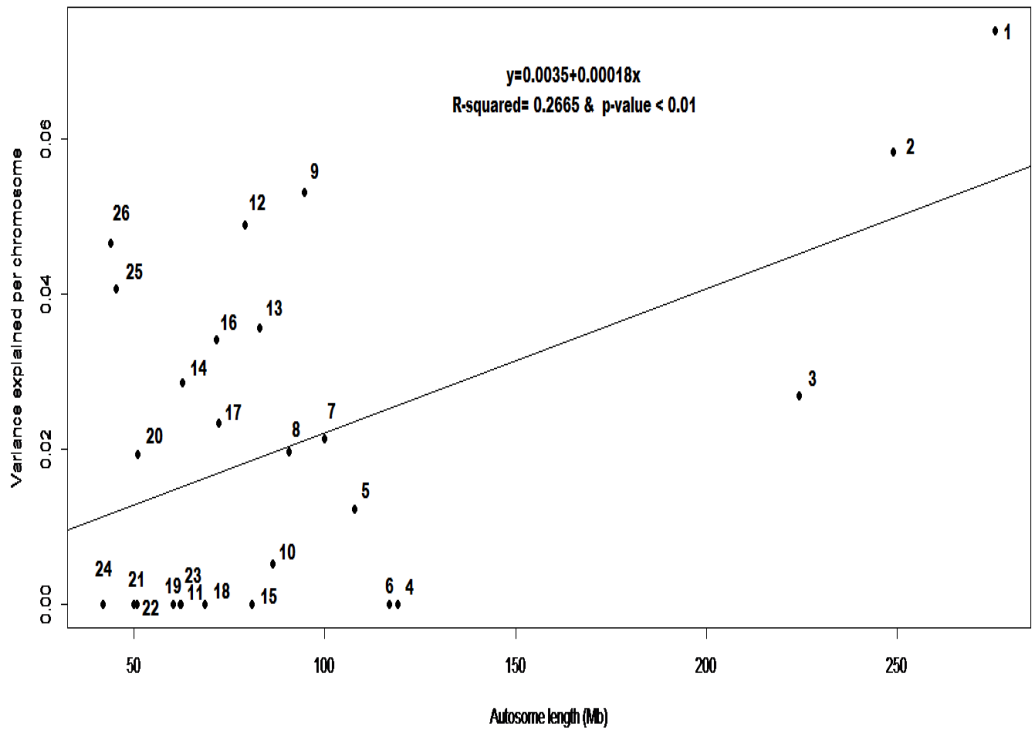

Figure 1. Amount of genomic variance explained per chromosome. The equation ($y = 0.0035 + 0.00018x$) corresponds to linear regression where y is the genomic variance explained by each chromosome and x is the chromosomal length in mega bases (Mb).