ESTIMATING THE GENETIC (CO)VARIANCE EXPLAINED PER CHROMOSOME FOR TWO GROWTH TRAITS USING A HALF SIB DATA STRUCTURE IN SHEEP

C. Esquivelzeta-Rabell[1], N. Moghaddar[1,2], S. Clark[1,2] and J.H.J. van der Werf[1,2]

[1]School of Environmental and Rural Science, University of New England, NSW, Australia
[2]Cooperative Research Centre for Sheep Industry Innovation, Armidale, NSW, Australia

SUMMARY

To detect how much genetic variance is accounted for by different genomic regions one first step is to work at the chromosomal level. We used a half sib data structure for two growth traits in sheep as a potentially powerful design to partition the genetic variance across chromosomes. Records for post weaning weight (PW) and scan C site back fat (CF) were used from 5,239 merino sheep. The model of analysis accounted for population structure by fitting genetic group effects as well as the numerator relationship matrix (A) or the first five principal components (PC). Different approximations were compared fitting the genomic relationship matrix (G) based on 48,599 markers, or on single nucleotide polymorphisms of an individual chromosome. The correlation between chromosome length (L) and variance explained per chromosome ($\sigma^2_{g_i}$) was 0.53 and 0.70 for PW and CF correspondingly, however significant differences in ($\sigma^2_{g_i}$/L) were found between chromosomes, ranging from 0% to 17.5%. Some chromosomes explained more variance and covariance than expected, under the assumption that it is proportional to the chromosome size; suggesting that some chromosomes clearly harbor more QTL. Some chromosomes show a covariance of opposite sign indicating they could be used in selection to 'break' an unfavourable correlation (e.g. chromosome 8). These results represent a powerful source of information for genomic selection.

INTRODUCTION

The tracking of chromosome segments through a pedigree is becoming more feasible due to the availability of abundant genomic information. However with the ever increasing density of genetic markers, there is also an increasing ambition to work out which variants actually are responsible for the observed quantitative genetic variation.

One first step to detect how much genetic variance is accounted for by different genomic regions is to work at the chromosomal level. Previous studies found a linear relationship between chromosome length and variance explained for human traits (Yang *et al.* 2010 and Visscher *et al.* 2007), production and fitness related traits in dairy cattle (Jensen *et al.* 2012) and production traits in sheep (Daetwyler *et al.* 2012).

The advantage of data on sheep populations is that the data structure is usually based on relatively large half sib families. This provides a powerful design for determining segregation based on linkage. The design is not suitable for LD mapping; hence the accuracy of mapping QTL positions would be low. However, the latter is less relevant for determining the amount of genetic variance explained per chromosome.

The objectives of this study were to estimate the genetic variance and covariance for two growth traits in sheep, determine the amount of additive genetic variance explained by each chromosome and to investigate the best model to correct for population and pedigree structure.

MATERIAL AND METHODS

Data for this study was obtained from the Information Nucleus program of the CRC for Sheep Industry Innovation. Details on this program and its design are described by van der Werf *et al.*

(2010). The data set comprised a total of 2,455 purebred merino lambs with phenotypes for two growth traits (post weaning weight: PW and scanned C site back fat: CF), pedigree and genotype data. The animals were descended from 139 sires and the associated pedigree file contained 10,559 animal identities from over 22 generations. The pedigree information was used to compute a numerator relationships matrix (A) for the animals with phenotypic records using the R package 'pedigree' (Coster 2012). Genotypic information consisted of SNP marker genotypes obtained using the OvineSNP50 BeadChip assay (Illumina, San Diego, USA). After quality control (Moghaddar *et al.* 2014) and imputing missing genotypes with Beagle 3.2 (Browning and Browning 2007), genotype information on 48,599 SNP was used to derive a genomic relationships matrix (G), scaling G to be analogous to A following VanRaden (2008).

The general model used to analyse the data was: $y = Xb + Z_a a + Z_m m + Z_q Qq + e$, where vector *b* included fixed effects of sex of lamb (ram: 1 or ewe: 2), birth type/rearing type (single: 1/1, twins: 2/2 or triplets: 3/3 and their combinations), management group, age of dam and weaning age; *a* is the random additive genetic effect of the lamb, *m* is the maternal permanent environmental effect and *q* is a genetic group effect. The genetic group consisted of merino strain (depending on the type of wool) where we regressed on strain proportion. Different models were explored to be able to partition the additive genetic variance into components that can be explained by markers while correcting for population structure, fitting into the model A and G individually (model 1 and 2 correspondingly) or simultaneously (model 3), together with genetic groups (merino strain) derived from a deep pedigree analysis, or fitting the first five principal components (PC) as a covariate (model 4). To decompose the variance components into 26 chromosomes, individual G_i were built based on marker information on the i^{th} chromosome and fitted simultaneously in the model (model 5), together with A (model 6) and with PC as a covariate (model 7). A bivariate analysis was also performed using the estimated variance components per trait resulting from the univariate analysis to define the starting values of the (co)variance matrices structures. The variance components for the first 4 models and the bivariate analysis were analysed using ASReml 3.0 software (Gilmour *et al.*, 2009). Models 5 to 7 were analysed using GCTA software (Yang *et al.*, 2011b).

RESULTS AND DISCUSSION

When pedigree (**A**) and marker based (**G**) relationship matrices were fitted individually we found that for PW SNPs capture additive genetic relationships among individuals and also effects of QTL (Table 1), G explaining more variation (36.80 %) than A (30.81%); however in the case of CF the results were opposite, A explained slightly more variation (25%) than G (22.61%). Nevertheless for both traits the log likelihood was higher when fitting only G in the model compared to fitting only A. Results from model 3, in which A and G were fitted simultaneously showed that most of the variance was partitioned toward G agreeing with previous reports (Haile-Mariam *et al.* 2013 and Jensen *et al.* 2012).

Model 4 was investigated as an alternative solution to correct for population structure avoiding co-linearity between the variance components. Results showed that variance explained by G after accounting for population structure using the first five PC was equal to 6.69 for PW and 0.09 for CF and the phenotypic variance was similar to the one calculated with the rest of the models.

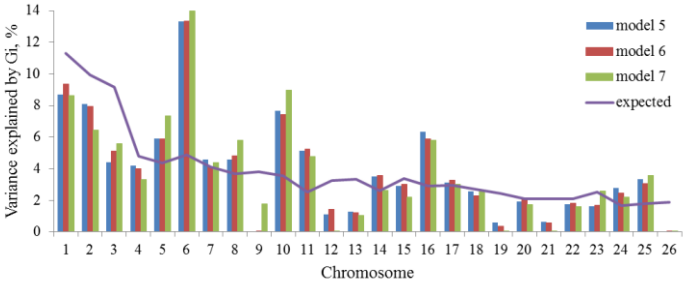
Genetic variance explained per chromosome using models 5 to 7 showed that the variance explained by each G_i varied and was somewhat related to the chromosome length (Figure 1) agreeing with Daetwyler *et al.* (2012). The correlation between variance explained and chromosome length was weaker for PW (0.55) than for CF (0.70) and marked differences in genomic variance explained were found for some chromosomes. For example, chromosome 6 explaining the higher amount of genomic variance for PW (13.3 to 13.7% depending on the model), followed by chromosome 1 (8.6 to 8.8%), 2 (6.4 to 8.2%), 10 (7.6 to 8.9%), 16 (5.8 to

6.4%) and chromosome 26 is estimated to contribute 0% variance.

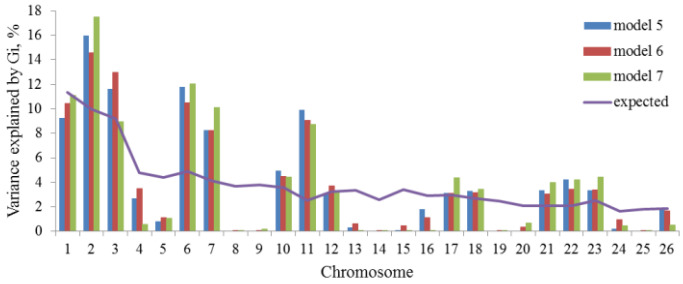
Table 1. Variance components estimates using different mixed linear models for post weaning weight (PW) and scanned C site back fat (CF).

PW	σ_a^2	σ_g^2	σ_m^2	σ_e^2	LogL
Model 1 (A)	6.74		2.21	12.92	-9485.64
Model 2 (G)		7.75	1.77	11.54	-9393.23
Model 3 (A+G)	1.33	7.74	1.44	10.88	-9393.19
Model 4 (G+PC)		6.69	2.38	12.12	-10358.2
CF					
Model 1 (A)	0.12			0.36	-1293.60
Model 2 (G)		0.11		0.37	-1279.29
Model 3 (A+G)	0.03	0.10		0.35	-1279.09
Model 4 (G+PC)		0.09		0.43	-1197.29

The difference between variance explained per chromosome and the expected variance explained, under the assumption that the genetic variance is proportional to the size of the chromosome (Daetwyler *et al.*, 2012), was also calculated (Figure 1) showing that some chromosomes explain more variance than expected; reflecting the relative QTL density on each chromosome, e.g. chromosome 6 for both traits with and extra variance of 9.1% for PW and 7.1% for CF; and others contribute with 0% of the variance explained, e.g. chromosomes 8, 9, 14, 19 and 25 for CF.



a)



b)

Figure 1. Genomic proportion of variance relative to the total size of the genome calculated per chromosome for post weaning weight (a) and scanned C site back fat (b) using models 5, 6 and 7. The line indicates the expected proportion of genomic variance explained per chromosome based on its size.

A bivariate model was used to estimate covariances and correlations per chromosome between growth traits in sheep. We found that 6 chromosomes have large covariance effects (Figure 2) and 5 have a negative covariance, e.g. chromosome 8, revealing that a small number of chromosomes appeared to 'break' the positive genetic correlation (overall genetic correlation ~ 0.5).

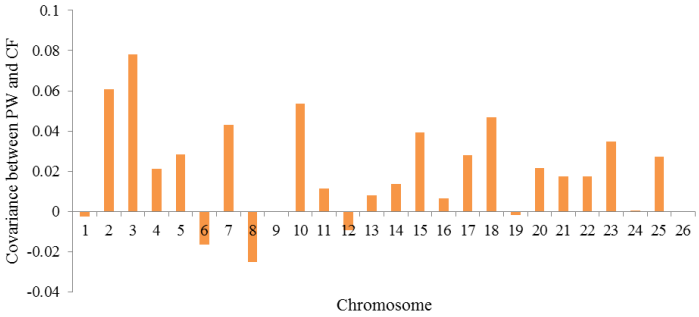


Figure 2. Covariance between post weaning weight (PW) and scanned C site back fat (CF) in merino lambs from bivariate analysis per chromosome.

The main conclusions from the present study are that the inclusion of PC in the model corrects for population structure avoiding co-linearity between the variance components. While the additive genetic variance explained per chromosome is partially related to chromosome length, considerable differences between chromosomes in the amount of additive genetic variance explained were found and a small number of chromosomes appeared to 'break' the positive genetic correlation.

The approach presented in our study provides relevant information to the understanding of the genetic underlying complex trait variation and represents a powerful source of information for genomic selection.

REFERENCES

Browning S.R. and Browning B.L. (2007) *Am. J. Hum. Genet.* **81**: 1084.

Coster A. (2012). pedigree: Pedigree functions. R package version 1.4. http://CRAN.R-project.org/package=pedigree

Daetwyler H.D., Kemper K.E., van der Werf J.H.J. and Hayes B.J. (2012) *J. Anim. Sci*. **90**: 3375.

Gilmour A.R., Gogel B.J., Cullis B.R. and Thompson R. (2009) ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Haile-Mariam M., Nieuwhof G.J., Beard K.T., Konstatinov K.V. and Hayes B.J. (2013) *J. Anim. Breed. Genet*. **130**: 20.

Jensen J., Guosheng S. and Per M. (2012) *BMC Genetics*. **13**: 44.

Moghaddar N., Swan A.A., van der Werf J.H.J. (2014) *GSE.* **46**: 58.

van der Werf J.H.J., Kinghorn B.P. and Banks R.G. (2010) *Anim. Prod. Sci*. **50**:998.

VanRaden P.M. (2008) *J. Dairy Sci*., **91**: 4414.

Visscher P.M., Macgregor S., Benyamin B., Zhu G., Gordon S. *et al.* (2007) *Am. J. Hum. Genet*. **81**: 1104.

Yang J., Benyamin B., McEvoy B., Gordon S., Henders A.K. *et al.* (2010) *Nat. Genet*. **42**: 565.

Yang J, Lee SH, Goddard ME and Visscher PM. (2011) *Am. J. Hum. Genet*. **88**(1): 76.