Genomic-Based Estimation of Genetic (Co)variances Applied to Mastitis Traits

L.P. Sørensen[*], T. Mark[†], L. Janss[*] and M.S. Lund[*]

Introduction

Genome-wide (GW) selection analysis based on SNP information (e.g. Meuwissen *et al*., 2001) has become widely used in breeding programs for dairy cattle, but also for broilers and pigs, because it reduces generation intervals and increases accuracies of predicted breeding values for a variety of traits. However, genomic models can also be used to study genome-wide effects on various phenotypic characteristics such as infectious diseases.

Mastitis is the most important disease in Danish dairy cattle (Krogh and Trinnerup, 2008). Treatments of mastitis and results from bacteriological analyses of milk samples from mastitic cows are recorded in the National Cattle Database. This kind of data provides a unique opportunity to study both genome-wide and chromosome-wise effects of mastitis in general, but also of pathogen-specific mastitis. With the use of a multi-trait GW model it is possible to estimate both genome-wide and chromosome-wise covariances among different mastitis traits and among mastitis traits and mastitis-related traits such as somatic cell count (SCC). Marker-based heritabilities have been estimated in humans using identical-by-descent relationship matrices (Visscher *et al.*, 2006). This approach is somewhat limited, and similar work has not been carried out in animals.

SNP-based GW models typically employ fixed prior parameters, which makes these models less suited for estimation of genetic (co)variances. The models that construct a genomic relationship matrix can be used to estimate (co)variances in a REML approach, but studying (co)variances per chromosome would be computationally prohibitive. Using multi-variate GW methodology for mastitis, it is possible to build a (co)variance matrix of allele substitution effects. In this, study we use a SNP-based GW model, which is extended to estimate hyper parameters in the prior distribution of allele substitution effects from the data. Thereby, the method can estimate genetic (co)variances while also being computationally feasible. Thus, the objectives of this project were to estimate heritabilities and genetic correlations for mastitis traits at genome and chromosome level using a GW approach.

Material and methods

Phenotypic data. Records of mastitis treatments, test-day SCC, and pathogen information from Holstein cows, which calved for the first time from 1998 to 2009, were extracted from

[*] Dept. of Genetics and Biotechnology, Aarhus University, DK-8830 Tjele, Denmark
[†] Dept. of Basic Animal and Veterinary Sciences, University of Copenhagen, DK-1870 Frederiksberg C, Denmark

the Danish National Cattle Database. The traits to be analyzed were defined as mastitis caused by *Staphylococcus aureus* (AUR), coagulase-negative staphylococci (CNS), *Escherichia coli* (COL), *Streptococcus dysgalactiae* (DYS) and *Streptococcus uberis* (UBE) and mastitis regardless of causative pathogen (MAS). The SCC trait was defined as lactation average SCC (LASCC) between 5 and 170 days after calving. After editing, the incidences of the mastitis traits were 0.026, 0.026, 0.021, 0.024, 0.029, and 0.22 for AUR, CNS, COL, DYS, UBE, and MAS, respectively. Progeny trait deviations (PTD) of the traits, similar to daughter yield deviations (VanRaden and Wiggans, 1991), were used as response variables in a GW model.

Marker data. The bulls used in this study were genotyped using Illumina Bovine SNP50 BeadChip (Illumina, San Diego, Ca). After editing, a total of 1844 bulls had 200,149 daughters with mastitis and LASCC data, and a total of 36,006 SNP were available.

Genomic model. A bivariate genomic model was used for analysis of PTDs with weighted residuals. This model was based on the Bayesian Variable Selection Method (George and McCulloch, 1993), using one "switch" per marker to select marker-effects on both traits either from a small-variance or a large-variance distribution, and extended to estimate residual variances, residual covariance, and the variances of the large-variance distributions for marker effects from the data. Currently, the model cannot accommodate polygenetic effects. This model assumed a prior zero correlation between the effects of a marker on the two traits, and the prior probability to select markers in the large-variance distributions was taken as 0.1. At the first level this bivariate model is:

$$\begin{cases} y_1 = \mathbf{1}\mu_1 + \sum_{i=1}^{M} \mathbf{x}_i b_{1i} + \nu_1 \mathbf{W}_1^{-1/2}\mathbf{l} + e_1 \\ y_2 = \mathbf{1}\mu_2 + \sum_{i=1}^{M} \mathbf{x}_i b_{2i} + \nu_2 \mathbf{W}_2^{-1/2}\mathbf{l} + e_2 \end{cases}$$

where y_1 and y_2 are vectors with PTDs for the two traits, μ_1 and μ_2 are PTD means of each trait, $\mathbf{x}_i$ are vectors of genotypes, b_{ki} is the random regression coefficient modeling the effect for SNP i on trait k, $\mathbf{W}$ is a diagonal matrix with weights ($1/SEP^2$ where SEP is standard error of predictions of the posterior PTD samples), $\mathbf{l}$ is a vector of latent effects that models residual covariance, ν_1 and ν_2 are scale factors for the effect of the latent vector $\mathbf{l}$ on each trait which can be interpreted as the elements of the first eigenvector of the residual variance-covariance matrix, and e_1 and e_2 are residuals.

Analyses were performed using MCMC methodology. Inferences were based on 50,000 samples with a burn-in of 10,000 samples. The primary samples of the model parameters were used to compute genetic and residual (co)variances per chromosome and subsequent marker-based, genome-wide heritabilities and genetic correlations were computed. Here, we only present results from bivariate analyses between AUR and the remaining 6 traits.

Results and discussion

Estimates of genome-wide heritabilities (Table 1) were generally much larger than the pedigree-based heritabilities (Sørensen *et al*., 2009a), and for AUR large variation was seen

depending on which trait AUR was analyzed with. However, the results may be obscured because PTDs were used instead of the raw phenotypes, which reduces residual variance and increases heritability. Genome-wide genetic correlations (Table 1) between AUR and the remaining traits were all positive but all much lower than pedigree-based genetic correlations (Sørensen *et al*., 2009b). For example the correlation between AUR and MAS was suspected to be much higher because a part-whole relationship exists between these two traits. The genetic correlations were expected to be comparable to pedigree-based genetic correlations. However, the results from the present study show that they are clearly underestimated. The reason for this is most likely because of the prior assumption of the model of zero covariance between SNP effects for the different traits. We are currently working on changing the model so correlations between markers (b_1, b_2) and residual (e_1, e_2) are estimated from data. This is expected to capture more of the genetic correlation between the traits.

Table 1: Estimates of SNP-based heritabilities (h^2) for mastitis traits and lactation average somatic cell count and genetic correlations (r_g) between *Staph. aureus* mastitis and the remaining traits

Traits	AUR	CNS	COL	DYS	UBE	MAS	LASCC
h^2	0.25-0.50	0.45	0.31	0.27	0.38	0.61	0.28
r_g	-	0.15	0.13	0.16	0.13	0.06	0.07

The novelty of the present approach is the ability to estimate chromosome-wise heritabilities and covariances. Heritability explained per chromosome was clearly proportional to chromosome size, i.e. larger chromosomes explained more variance. This pattern was seen for all the investigated traits. Figure 1a shows an example of chromosome-wise heritability for AUR. For this trait, bovine autosomes (BTA) 1, 11, 16, 19, and 23 explained more genetic variance than could be expected relative to their size. In particular, BTA 19 was interesting as this chromosome explained more variance than expected for all the mastitis traits, whereas for LASCC, BTA 6 and 20 seemed to be of interest. These results may be used to point of chromosomes harboring important QTL affecting the traits, both unspecific and pathogen-specific.

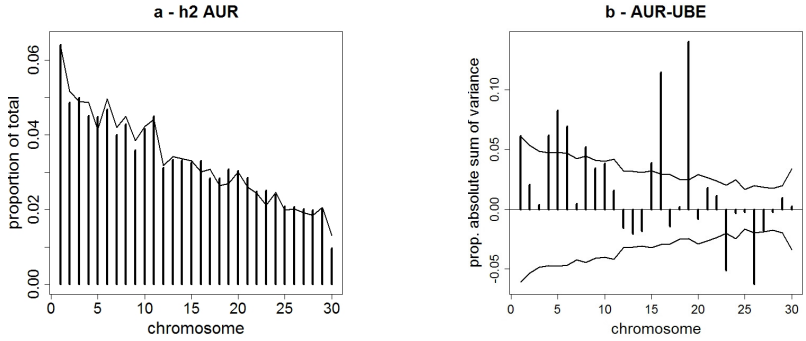

Figure 1: Examples of chromosome-wise heritability (a) and chromosome-wise covariance (b) given as proportion of the absolute sum of covariance for each chromosome.

Bars reaching above the solid line indicate that the chromosome explains more variance than can be expected (solid line) from the relative size of the chromosome.

Most chromosome-wise covariances among the investigated traits were positive which also was reflected in the genome-wide correlations. This indicates that genes controlling immune response to one pathogen also control immune response to other pathogens. As shown in figure 1b, large negative covariances (BTA 23 and 26) also exist between the traits. However, in this case the genome-wide covariance was positive due to several chromosomes showing positive covariance (e.g. BTA 16 and 19). Relationships between AUR and LASCC showed similar patterns (results not shown).

Conclusion

A genomic model was used to estimate "genomic" variances and covariances based on estimating individual marker effects. This allows for separation of variance and covariances according to chromosomes or genomic regions, which would be virtually impossible when using marker-based genomic relationship matrices. From other work, the genomic variances and heritabilities are known to be similar to pedigree-based estimates. However, in the model used here, covariances appear under-estimated and further development is in progress to improve this. This approach could be suitable for estimation of heritabilities in populations where pedigrees may be incomplete or for estimation of disease-susceptible regions.

Acknowledgement

This work was performed in the project 'Genomic Selection – from function to efficient utilization in cattle breeding (grant no. 3412-08-02253) funded by the Danish Directorate for Food, Fisheries and Agri Business, VikingGenetics, Nordic Genetic Evaluation, and Aarhus University.

References

George, E.I. and McCulloch, R.E. (1993). *J. Am. Stat. Assoc*. 88:881-889.

Krogh, K. and Trinderup, M. (2008). http://www.lr.dk/kvaeg/informationsserier/kvaegforsk/1902.htm.

Meuwissen, T.H.E., Hayes, B.J., Goddard, M.E. (2001). *Genetics* 157:1819-1829.

Sørensen, L.P., Madsen, P., Mark, T. *et al*. (2009a). *Animal* 3:647-656.

Sørensen, L.P., Madsen, P., Mark, T. *et al*. (2009b). *J. Dairy Sci*. 92:3457-3471.

VanRaden, P.M. and Wiggans, G.R. (1991). *J. Dairy Sci*. 74:2737-2747.

Visscher, P.M., Medland, S.E., Ferreira, M.A.R. *et al.* (2006) *PLoS Genet*. 2:e41