Estimation Of Heritability For Dairy Traits, Combining Pedigree With Dense SNPs Information On Some Animals

R.F. Veerkamp[*], H.A. Mulder[*] and M.P.L. Calus[*]

Introduction

Genotyping animals with dense SNP information across the genome is becoming cheaper every year. In practice there has been a lot of attention to use this information for breeding value estimation (Hayes et al. (2009)). Developments were on models allowing estimation of individual variation for an IBD haplotype per locus, models estimating effects per SNP (Calus et al. (2008)), and models that use SNP information to form a genomic relationship matrix (G-RM) that can replace the numerator relationship matrix in routine BLUP evaluations (VanRaden (2008)). The G-RM represents the true relationship between relatives more precise than the numerator relationship based on pedigree information, because it reflects that relationships may deviate from the expected average relationship due to Mendelian sampling. Two disadvantages of the G-RM are that information on known effects of individual SNPs is ignored, and that non genotyped animals can not be included easily. To overcome the latter disadvantages (Aguilar et al. (2010)) developed a method to combine the genomic and numerator relationship matrices for genotyped and non-genotyped animals. To overcome the first disadvantage we use a G-RM weighted by the size of individual SNP effects, based on a suggestion from Goddard (2009) to parameterize mixed model equations taking estimated SNP effects into account.

The objective of this study was to investigate how dense marker information can be used to improve estimation of heritabilities for dairy traits that are scarcely recorded on some animals, i.e. feed intake and live weight. Firstly by comparing estimated variance component and their standard errors for three different analyses: i) on 639 animals with pedigree information; ii) a subset of 517 genotyped animals and iii) a combined analysis of genotyped and non genotyped animals. Secondly by weighting individual SNP effects in the G-RM.

Material and methods

Description of data. Data on 639 Holstein-Friesian heifers born between 1990 and 1997 were collected during the first 15 wk of lactation. All cows were fed *ad libitum*. Live weight, milk yield and milk composition were recorded weekly, and feed intake was recorded daily using automated feed intake units, More comprehensive details on the data used can be found in Veerkamp et al. (2000). A subset of these animals had DNA available and these were genotyped using the Illumina 50K SNP panel (54,001 SNP in total). Quality control checks included a call rate for each SNP of over 90%, a GenCall score >0.2 and a GenTrain score >0.55, a minor allele frequency of >2.5% and a lack of deviation from Hardy Weinberg

[*] Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, PO Box 65, 8200 AB Lelystad.

equilibrium (for more details on the editing see Verbyla et al. (2010)). After all editing steps, in total, 43,011 SNP and 517 animals were retained that had genotypes and phenotypes for all yield, intake and live weight traits considered here.

Pedigree and genomic relationships. The numerator relationship matrix (A-RM) was set up for the 639 animal with data and 3363 ancestors (Meuwissen and Luo (1992)). The genomic relationship matrix (G-RM) was set up for the 517 genotyped following (VanRaden (2008)):

$$GRM = \frac{\mathbf{ZZ}'}{2\sum p_i(1-p_i)}$$

where **Z** is a matrix that relates SNP alleles to individuals and p_i is the second allele frequency of SNP_i. The A^{-1} (full pedigree), A^{-1} (pedigree genotyped animals), and G^{-1} were combined to H^{-1} following (Aguilar et al.)).

SNP weighted genomic relationships. To take account of different weights for SNP in the G-RM, as a first step the SNP effects were estimated using the Gibbs sampler described earlier (Meuwissen and Goddard (2004); Calus et al. (2008); Calus et al. (2009)) with the following model: $y_i = \mu + fixed\_effects + \sum_{j=1}^{nloc}\sum_{k=1}^{2} SNP_{ijk} + e_i$,

where y_i is the phenotypic record of animal *i*; μ is the average phenotypic performance; SNP_{ijk} is is a random effect for the kth (k=1,2) SNP allele *j* (of nloc loci) of animal *i*, and e_i is a random residual for animal *i*. The parameterisation assumed the SNP effects came from two distributions (i.e. BayesC). The weighted genomic relationship matrix (WG-RM) was now calculated using the estimated allele substitution effects at a locus, with the formula

$$WGRM = \frac{ZDZ'}{2\sum p_i(1-p_i)}$$

where D is a diagonal matrix with weights per SNP on the diagonal. The weights were proportional to the squared allele substitution effects and rescaled to be 1.0 on average across all loci. Note that replacing D by I yields the G-RM as a result. WG-RM matrices were calculated weighting all SNP by their squared estimated allele substitution effect (ALL). Alternatively only SNPs with a threshold value for the posterior probability of the SNP being associated to a QTL received a weight proportional to their squared estimated allele substitution effect. In this scenario, the weights of all SNPs with a posterior probability below the threshold were averaged within this group of SNPs. Finally, all weights were scaled across loci to be 1.0 on average.

Statistical analyses. To estimate variance components for the random animal effect associated with either of these relationship matrices ASREML (Gilmour et al. (2000)) was used with the following general model: $y_{ij} = \mu + Fixed\_effects_j + a_i + e_{ij}$ where y_{ij} is the phenotype (milk, fat and protein yield or composition, dry matter intake or live weight) of animal i; fixed_effects are the effects: year-quarter (25 levels) and age at calving (3 levels); a_i is the random additive genetic effect of the i_{th} animal; and e_{ij} are the residuals. $Var(a)=A\sigma_a$ or $G\sigma_a$ or $H\sigma_a$.

Results and discussion

Estimated variance components of the models are given in Table 1. Generally the estimates based on pedigree relationships (A-RM) gave higher genetic variances than the estimates based on genomic information. For all models standard errors (SE) are large, as expected for such a small dataset. When the pedigree was used with phenotypes from the genotyped animals only, standard errors were highest. Additional phenotypes did not improve the SE as much as when using the G-RM, despite using 112 records less. The combined analysis (H-RM) uses all phenotypes and the more precise genomic relationships for the genotyped animals, and therefore gives the most precise estimate of the heritability for all traits.

Table 1: Estimates of genetic variance, heritability (h^2) and standard error (SE) of h^2, using numerator (A-RM), genomic (G-RM) or combined (H-RM) relationship matrix, and 639 or 517 phenotypes depending on the model used.

	σ_a				h^2				SE			
RM	**A**	**A**	**G**	**H**	**A**	**A**	**G**	**H**	**A**	**A**	**G**	**H**
# phenotypes	517	639	517	639	517	639	517	639	517	639	517	639
Milk (kg/d)	8.6	7.8	7.6	7.2	0.48	0.44	0.43	0.41	0.13	0.11	0.10	0.09
Fat (kg/d)	1.37	1.22	1.20	1.12	0.48	0.44	0.43	0.42	0.13	0.11	0.10	0.09
Protein (kg/d)	0.55	0.51	0.44	0.43	0.41	0.39	0.33	0.33	0.13	0.11	0.10	0.09
Fat (%)	0.150	0.143	0.123	0.130	0.89	0.84	0.77	0.79	0.10	0.08	0.08	0.07
Protein (%)	0.037	0.032	0.030	0.028	0.81	0.71	0.70	0.65	0.10	0.10	0.08	0.08
DMI (kg/d)	2.5	2.3	1.8	1.9	0.83	0.77	0.66	0.67	0.11	0.10	0.09	0.08
LW (kg)	744	841	678	748	0.50	0.56	0.46	0.51	0.13	0.11	0.10	0.09

The presumption was that when the genomic relationship was augmented by the individual SNP effects, estimates of the heritability become more precise than when weighting all SNPs equally. This decrease in the standard error of het heritability was observed (Table 2). For fat% when the three SNPs were selected with the highest posterior probability of being linked to a QTL, the standard error dropped from 0.08 to 0.07. One of these SNPs was linked to the DGAT gene with a large effect on fat%. Weighting more SNP reduce the SE even further, but weighting all SNP effects according to the estimated allele effects did not work (Table 2). Phenotypic variances became unrealistic large for all traits and residual variances became close to zero. This was unexpected because SNP effects did not explain the full phenotypic variance. A possible explanation might be that the there is an autocorrelation between the relationship matrix and phenotypic values, because they were estimated from the same data.

Table 2: Estimates of variances (G = genetic, R = residual, P=phenotypic), heritability (h^2) for fat% and standard error (SE) using WG-RM genomic relationship matrix with higher weights for some (# SNP) loci according to the estimated allele effects, for all SNP, or only SNP that had posterior probability >0.001, >0.01, >0.05, >0.10.

RM Post. Prob.	G none	WG p>0.10	WG p>0.05	WG p>0.01	WG p>0.001	WG All
# SNP	-	3	7	116	4237	43011
G	0.12	0.12	0.13	0.13	0.14	0.61
R	0.04	0.05	0.04	0.04	0.04	0.00
P	0.16	0.17	0.17	0.17	0.18	0.61
h^2	0.77	0.72	0.75	0.77	0.79	1.00
SE	0.08	0.07	0.07	0.06	0.06	0.00

Conclusion

Using the genomic relationship matrix (based on 43,011 SNP) improved the estimation of the heritability, even when 517 phenotypes were used instead of 639 phenotypes with pedigree relationships. Combining both pedigree and genomic relationships gave the most precise estimate of the heritability.

Acknowledgements

This work was carried out as part of the RobustMilk project that is financially supported by the European Commission under the Seventh Research Framework Programme, Grant Agreement KBBE-211708. The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the Commission or its services.

References

Aguilar, I., Misztal, I., Johnson, D.L. *et al.* (2010). *J. Dairy Sci.*, 93:743-752.

Calus, M., Meuwissen, T., Windig, J. *et al.* (2009). *Genet. Sel. Evol.*, 41:11.

Calus, M.P.L., Meuwissen, T.H.E., De Roos, A.P.W. *et al.* (2008). *Genetics*, 178:553- 561.

Gilmour, A.R., Cullis, B.R., Welham, S.J. *et al.* (2000). NSW Agriculture, Orange Agricultural Institute, Forest Road, Orange, NSW, 2800, Australia.

Goddard, M. (2009). Genetica, 136:245-257

Hayes, B.J., Bowman, P.J., Chamberlain, A.J. *et al.* (2009). *J. Dairy Sci.*, 92:433-443.

Meuwissen, T.H.E. and Goddard, M.E. (2004). *Genet. Sel. Evol.*, 36:261 - 279.

Meuwissen, T.H.E. and Luo, Z. (1992). *Genet. Sel. Evol.*, 24:305-313.

VanRaden, P.M. (2008). *J. Dairy Sci.*, 91:4414-4423.

Veerkamp, R.F., Oldenbroek, J.K., van der Gaast, H.J. *et al.* (2000). *J. Dairy Sci.*, 83:577.

Verbyla, K.L., Calus, M.P.L., Mulder, H.A. *et al.* (2010). *J. Dairy Sci., 93:2757-2764.*