

Rothamsted Repository Download

A - Papers appearing in refereed journals

Veerkamp, R. F., Mulder, H. A., Thompson, R. and Calus, M. P. L. 2011. Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped. *Journal of Dairy Science*. 94 (8), pp. 4189-4197.

The publisher's version can be accessed at:

- <https://dx.doi.org/10.3168/jds.2011-4223>

The output can be accessed at: <https://repository.rothamsted.ac.uk/item/8q97v>.

© American Dairy Science Association®, 2011



Genomic and pedigree-based genetic parameters for scarcely recorded traits when some animals are genotyped¹

R. F. Veerkamp,^{*2} H. A. Mulder,^{*} R. Thompson,[†] and M. P. L. Calus^{*}

^{*}Animal Breeding and Genomics Centre, Wageningen UR Livestock Research, 8200 AB Lelystad, the Netherlands

[†]Biomathematics Unit, Rothamsted Research Ltd., Harpenden AL5 2JQ, United Kingdom

ABSTRACT

Genetic parameters were estimated using relationships between animals that were based either on pedigree, 43,011 single nucleotide polymorphisms, or a combination of these, considering genotyped and non-genotyped animals. The standard error of the estimates and a parametric bootstrapping procedure was used to investigate sampling properties of the estimated variance components. The data set contained milk yield, dry matter intake and body weight for 517 first-lactation heifers with genotypes and phenotypes, and another 112 heifers with phenotypes only. Multivariate models were fitted using the different relationships in ASReml software. Estimates of genetic variance were lower based on genomic relationships than using pedigree relationships. Genetic variances from genomic and pedigree relationships were, however, not directly comparable because they apply to different base populations. Standard errors indicated that using the genomic relationships gave more accurate estimates of heritability but equally accurate estimates of genetic correlation. However, the estimates of standard errors were affected by the differences in scale between the 2 relationship matrices, causing differences in values of the genetic parameters. The bootstrapping results (with genetic parameters at the same level), confirmed that both heritability and genetic correlations were estimated more accurately with genomic relationships in comparison with using the pedigree relationships. Animals without genotype were included in the analysis by merging genomic and pedigree relationships. This allowed all phenotypes to be used, including those from non-genotyped animals. This combination of genomic and pedigree relationships gave the most accurate estimates of genetic variance. When a small data set is available it might be more advantageous for the estimation of genetic parameters

to genotype existing animals, rather than collecting more phenotypes.

Key words: genomic selection, heritability, genetic correlation, feed intake

INTRODUCTION

Genotyping animals with dense SNP information across the genome is becoming cheaper every year and a lot of attention has focused on using high-density SNP information for breeding value estimation (Hayes et al., 2009). Models have been developed to estimate the variation explained by identical-by-descent haplotypes or individual SNP (Calus et al., 2008). Other models for breeding value estimation have been suggested using the high-density SNP information to form a genomic relationship matrix (**GRM**), with the advantage that the GRM can replace the numerator relationship matrix based on pedigree information (**ARM**) in routine BLUP evaluations (VanRaden, 2008). Compared with the ARM, the advantage of the GRM is that the relationship between relatives are described more precisely, because it reflects that actual relationships may deviate from their expectation because of Mendelian sampling. Because the GRM reflects relationships between animals more precisely than the ARM, it can be hypothesized that, compared with estimates with the ARM, genetic variance components can be estimated more accurately with the GRM. If so, genotyping animals might become especially appealing when small data sets are available for genetic studies, for example for scarcely recorded traits that are so difficult or expensive to record that only few phenotypes are available (e.g., feed intake and BW of dairy cows).

When GRM models are used on scarcely recorded traits, a drawback might be that not all animals with phenotypes have DNA available. Therefore, using the GRM results in a loss of phenotypic records (i.e., records of those animals that have no DNA available). This is a common scenario where historical phenotypes are used to estimate genetic parameters. For BLUP evaluations, a unified method was developed to combine the ARM and GRM for genotyped and non-genotyped animals

Received January 30, 2011.

Accepted March 29, 2011.

¹The content of this paper is the sole responsibility of the authors, and it does not necessarily represent the views of the European Commission or its services.

²Corresponding author: Roel.Veerkamp@wur.nl

(Aguilar et al., 2010; Legarra and Misztal, 2008). This method might also be appealing for variance component estimation, because data of the genotyped and non-genotyped animals can be combined.

The objective of this study was to estimate variance components for milk yield, feed intake, and BW in a small data set using genetic relationships based on SNP, pedigree information, or combining both, and to compare the accuracy of the estimated genetic parameters across methods. A parametric bootstrap procedure was developed to obtain the sampling variance of the estimated values.

MATERIALS AND METHODS

Phenotypes and Genotypes

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. Milk yield (MY) and BW were recorded weekly, and DMI was recorded daily using automated feed intake units. Comprehensive details on the data used can be found in Veerkamp et al. (2000). A subset of these animals (603) had blood samples available for DNA extraction, and these were genotyped using the Illumina 50K SNP panel (54,001 SNP in total; Illumina Inc., San Diego, CA). Quality control checks included a call rate for each SNP of over 90%, a GenCall score >0.2, a GenTrain score >0.55, a minor allele frequency of >2.5%, and a lack of deviation from Hardy-Weinberg equilibrium (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 traits considered here.

Genetic Relationship Matrices

Pedigree Relationships. Assuming the infinitesimal model, the pedigree relationships were used to calculate the numerator relationship matrices (Meuwissen and Luo, 1992) for the 517 animals that had genotypes plus phenotypes available and the 639 that had phenotypes available, hereafter referred to as **A517** and **A639**, respectively. Pedigree was traced back for as many generations as available in the Dutch herd book (more than 4 for most animals). The 517 animals that were genotyped had 93 sires and 381 dams, and the complete set of 639 animals with phenotypes had 105 sires and 455 dams. In the full data set, one sire had at maximum 36 offspring with data, and 27 sires had one offspring only. More than 124 dams had 2 or more offspring with data available (14 dams had between 4 and 8 offspring) and 135 dams had also phenotypes themselves. Furthermore, 65 full-sibling groups had 2

to 5 siblings. These sibling groups occurred because of the use of embryo transfer in the herd.

Genomic Relationship. Assuming a genetic model in which all 43,011 SNP affect a trait with equal variance per SNP, a relationship matrix can be calculated based on SNP genotypes. This genomic relationship matrix (**G517**) was set up for the 517 genotyped animals following VanRaden (2008):

$$\mathbf{G} = \frac{\mathbf{Z}\mathbf{Z}'}{2\sum p_i(1-p_i)},$$

where p_i was the frequency of the second allele at locus i , and the elements of **Z** were derived by subtracting 2 times the allele frequency expressed as a difference of 0.5, that is $2(p_i - 0.5)$, from matrix **M** that specifies the 3 marker genotypes for each individual as -1, 0, or 1 (VanRaden, 2008). This **G** estimates the relationships between animals based on SNP information, and allows the relationship to deviate from the average pedigree relationship. However, the complete genome sequence is not available and the estimated genomic relationships are, therefore, subject to sampling error (Powell et al., 2010). Therefore, the properties of **G** were adjusted, considering that **G** is estimated with some error, following the procedure suggested by Yang et al. (2010):

$$\hat{\mathbf{G}} = \mathbf{G} + \mathbf{E} = \mathbf{A} + (\mathbf{G} - \mathbf{A}) + \mathbf{E},$$

where **E** is a matrix with estimation errors for $\hat{\mathbf{G}}$ and **A** is the matrix with the pedigree relationships.

Further, consider that the variances of those matrices are

$$V(\hat{\mathbf{G}} - \mathbf{A}) + V(\mathbf{G} - \mathbf{A}) + V(\mathbf{E}),$$

where $V(\mathbf{E}) = 1/N$, and N is the number of SNP used to calculate **G**.

The difference between **G** and **A** ($\hat{\mathbf{G}} - \mathbf{A}$) was then regressed back toward **A**, to account for the sampling variance in $\hat{\mathbf{G}}$. The adjusted **G** ($\tilde{\mathbf{G}}$) was calculated as follows:

$$\begin{aligned} \tilde{\mathbf{G}} &= \mathbf{A} + (\hat{\mathbf{G}} - \mathbf{A}) \times V(\mathbf{G} - \mathbf{A}) / [V(\mathbf{G} - \mathbf{A}) + V(\mathbf{E})] \\ &= \mathbf{A} + (\hat{\mathbf{G}} - \mathbf{A}) \times [V(\hat{\mathbf{G}} - \mathbf{A}) - 1/N] / V(\hat{\mathbf{G}} - \mathbf{A}). \end{aligned}$$

To account for the fact that the sampling error in relationships in $\hat{\mathbf{G}}$ depends on the value of the relationship, regression coefficients were separately calculated for bins of the relationships in **A**, being 0 to 0.10, 0.11 to 0.25, 0.26 to 0.50, and >0.50. In the last bin, parent-offspring pairs were not included, because the expectation of all those relationships is 0.5. Self-relationships

(i.e., diagonals of \hat{G}), were not regressed. The regression coefficients for the 4 bins were 0.978, 0.981, 0.990, and 0.993, respectively. The final relationship matrix \tilde{G} (**G517**) was used in the analysis.

Combined Genetic and Genomic Relationships. To be able to use phenotypic records on genotyped and non-genotyped animals, the pedigree and genomic relationships needed to be combined in one relationship matrix. This combined inverse relationship matrix for the 639 (**H639**⁻¹) animals was created using the inverses of **G517**, **A517**, and **A639** (Legarra and Misztal, 2008; Aguilar et al., 2010):

$$\mathbf{H639}^{-1} = \mathbf{A639}^{-1} + \begin{bmatrix} 0 & 0 \\ 0 & \mathbf{G517}^{-1} - \mathbf{A517}^{-1} \end{bmatrix}.$$

Statistical Model

Assuming the underlying genetic models, variance components were estimated with ASReml software (Gilmour et al., 2000) using the general model:

$$y_{ij} + \mu + fixed_effects_j + a_i + e_{ij},$$

where y_{ij} was the phenotype of animal i ; *fixed_effects* were year-quarter (25 levels) and age at calving (3 levels); a_i was the random additive genetic effect of the i th animal; $\text{Var}(a) = \mathbf{A}_{517}\sigma^2_a$ or $\mathbf{A}_{639}\sigma^2_a$, or $\mathbf{G}_{517}\sigma^2_a$, or $\mathbf{H}_{639}\sigma^2_a$; and e_{ij} were the random residuals. This general model was used to analyze MY, DMI, and BW together with separate fixed effects for each trait, and including covariances between the residuals for the 3 traits, and the estimated genetic effects. Standard errors of the estimated heritabilities and genetic correlations were approximated with ASReml using the (co-)variances of the estimated (co-)variances.

Parametric Bootstrapping

The accuracy of the estimated variance components can be inferred from the standard errors approximated by ASReml. However, the approximation might be less precise for small data sets and affected by the level of the estimated variances. Therefore, the sampling variance and standard errors of the genetic parameters were also obtained with a parametric bootstrapping procedure, using the same values to simulate data with all 4 relationship matrices (**A517**, **G517**, **A639**, and **H639**). Second, to investigate the effect of **G** and **A** on the values of the variances, data simulated with **A517** were analyzed with **G517** and vice versa. The first step was to create 250 replicated data sets for the 517 and 639 animals, respectively. These phenotypes

for each replicate were created by sampling (and adding) the genetic and residual effect for each animal from a multivariate normal distribution using the rmvnorm procedure in R (Genz et al., 2010). The (co-)variances between the genetic effects across traits and across relationships were calculated as the Kronecker product between the genetic covariances for the 3 traits and the relevant relationship matrix (**A517**, **A639**, **G517**, or **H639**). For the residual covariances, estimates were based on the Kronecker product between an identity matrix and the residual covariances between the traits. The phenotypes in each replicate were sampled from the respective matrices:

$$Y_{1\dots n} = MVN \left(\begin{bmatrix} a_{1,1} & \dots & \text{sym.} \\ \dots & \dots & \dots \\ a_{n,1} & \dots & a_{n,n} \end{bmatrix} \otimes \begin{bmatrix} \mathbf{A}_{1,1} & & \text{sym.} \\ \mathbf{A}_{2,1} & \mathbf{A}_{2,2} & \\ \mathbf{A}_{3,1} & \mathbf{A}_{3,2} & \mathbf{A}_{3,3} \end{bmatrix} \right) + MVN \left(\mathbf{I} \otimes \begin{bmatrix} \mathbf{R}_{1,1} & & \text{sym.} \\ \mathbf{R}_{2,1} & \mathbf{R}_{2,2} & \\ \mathbf{R}_{3,1} & \mathbf{R}_{3,2} & \mathbf{R}_{3,3} \end{bmatrix} \right),$$

where $Y_{i\dots n}$ are the sampled phenotypes for the traits MY, DMI, and BW for $n = 517$ or 639 animals; $MVN(V)$ samples from the multivariate normal distribution with mean zero and (co-)variance between the effects V ; $\mathbf{A}_{1\dots 3,1\dots 3}$ and $\mathbf{R}_{1\dots 3,1\dots 3}$ are the genetic and residual covariance matrix for MY, DMI, and BW, respectively (Table 1); $a_{1\dots n,1\dots n}$ are the genetic relationships between n animals coming from **A517**, **A639**, **G517**, or **H639**; the values taken for **A** and **R** were estimates from the **H639**-matrix analysis; and **I** is an identity matrix of size 517 (or 639). Variance components were estimated for each replicate with ASReml, using the model described earlier. Fixed effects were not simulated, although they have an effect on the stratification of the data and, therefore, the sampling properties of the genetic correlations. For this reason, fixed effects were included in the estimation model, albeit the expectation for each effect was zero. Standard errors for the heritability and genetic correlations were obtained by taking the standard deviation of the 250 replicates.

RESULTS

Variance Components for Milk, DMI, and BW

Genetic and phenotypic variances were estimated using numerator (**A517** and **A639**), genomic (**G517**) and combined (**H639**) relationship matrix (Table 2). The models with the SNP-based relationships gave lower genetic variances and heritability in comparison

Table 1. Heritability, genetic and phenotypic covariances (below diagonal), genetic and phenotypic variances (underlined on diagonal), and genetic and phenotypic correlations (above diagonal) used in the simulations

	Genetic			h^2	Phenotypic		
	Milk	DMI	BW		Milk	DMI	BW
Milk (kg/d)	<u>7.3</u>	0.36	0.09	0.42	<u>17.4</u>	0.36	-0.03
DMI (kg/d)	1.3	<u>1.8</u>	0.70	0.65	2.5	<u>2.8</u>	0.52
BW (kg)	6.7	25.7	<u>743</u>	0.51	-5.1	33.0	<u>1.466</u>

with the models that included the pedigree-based relationships. Although these deviations were within the standard errors of the estimates, the same data were used (e.g., **A517** and **G517**) and differences were larger than expected, especially for DMI. The models with the GRM gave a better likelihood (log-likelihood of **A517** and **G517** analyses were -3,392.4 and -3,382.7, respectively, and for **A639** and **H639** they were -4,206.4 and -4,196.2, respectively), suggesting a better fit of the data in comparison with the models with ARM.

For all models, standard errors of the heritability estimates were large (between 0.08 and 0.12), as expected for such a small data set. When the pedigree was used with phenotypes from the genotyped animals only (**A517**) standard errors were highest (0.12 for milk yield). The additional 122 phenotypes did improve the standard error to 0.11 (**A639**), but not as much as when using the GRM on the smaller data set (**G517**). This model gave for milk a standard error of 0.10, despite using 122 records fewer than the model **A639**. The combined analysis (**H639**) used all phenotypes and the more precise genomic relationships for the genotyped animals, and therefore, gave the most precise estimate of the heritability for all traits (standard error of 0.09 for milk).

Genetic correlations were closer to zero using the SNP-based relationships in the model, compared with using the pedigree-based relationship model (Table 3). However, little advantage appeared in terms of precision of the estimated genetic correlations, as standard

errors were similar between models with pedigree or SNP-based relationships. Using the additional 122 records decreased the standard errors for genetic correlations between all pairs of traits, as expected.

Bootstrapping

Across the replicates, mean estimates of the genetic parameters and variances (Table 4) were equivalent between the 4 models compared with the simulated values (Table 1). This is not surprising, because the data were simulated with the same model used in the analysis and the same simulated values were used in all 4 models. When the data simulated with **A517** were analyzed with **G517**, a considerable decrease (e.g., from 7.3 to 4.9 for milk) was observed for the genetic variance (Table 5). The reason may be that a considerable part of the genetic covariances between animals (based on GRM) does not correspond to covariances in the data simulated with ARM; for example, animals unrelated by pedigree might still be related through the GRM. Also, the average Mendelian sampling when simulated with **A517** is different than the Mendelian sampling in **G517**. When the data simulated with **G517** were analyzed with **A517**, estimates for the genetic variance were closer but slightly higher than the simulated values.

The distribution of the variance component values can be used to evaluate the accuracy of the different procedures, as demonstrated for the genetic correla-

Table 2. Estimates of genetic variance (σ_a^2), phenotypic variance (σ_p^2), and heritability (h^2) with their approximate standard errors obtained from ASReml software, using numerator (**A517** and **A639**), genomic (**G517**), and combined (**H639**) relationship matrices

	A517						G517					
	σ_a^2	SE	σ_p^2	SE	h^2	SE	σ_a^2	SE	σ_p^2	SE	h^2	SE
Milk (kg/d)	9.3	2.7	18.3	1.4	0.50	0.12	7.8	2.0	17.7	1.3	0.44	0.10
DMI (kg/d)	2.5	0.5	3.0	0.2	0.82	0.11	1.8	0.4	2.8	0.2	0.65	0.09
BW (kg)	844	224	1,527	116	0.55	0.12	688	169	1,468	106	0.47	0.10
	A639						H639					
Milk (kg/d)	8.4	2.2	17.9	1.2	0.47	0.11	7.3	1.8	17.4	1.1	0.42	0.09
DMI (kg/d)	2.3	0.4	2.9	0.2	0.78	0.10	1.8	0.3	2.8	0.2	0.65	0.08
BW (kg)	893	197	1,520	105	0.59	0.10	730	155	1,463	96	0.50	0.09

Table 3. Estimates of genetic correlations (below diagonal) and their approximate standard errors obtained from ASReml software (above diagonal) between milk, DMI, and BW using numerator (**A517** and **A639**), genomic (**G517**), and combined (**H639**) relationship matrices

	A517			G517		
	Milk	DMI	BW	Milk	DMI	BW
Milk		0.14	0.19		0.14	0.18
DMI	0.44		0.09	0.40		0.09
BW	0.37	0.78		0.25	0.75	
	A639			H639		
	Milk	DMI	BW	Milk	DMI	BW
Milk		0.13	0.17		0.13	0.16
DMI	0.44		0.08	0.35		0.09
BW	0.19	0.75		0.10	0.72	

tion between BW and milk yield (Figure 1). For all traits, standard deviations of the 250 estimates of the heritability (Table 6) were very close to the estimates of the standard error that ASReml gave (Table 2). For the models with the numerator relationship matrix, the sampling procedure gave 0.01 higher estimates for the standard deviation of the heritability than the estimates for the standard error given by ASReml. For the 2 models that include SNP relationships, hardly any difference existed between these estimates. Although not found for the standard error of the genetic correlations from ASReml, with the bootstrapping procedure a clear improvement of the accuracy of the estimated genetic correlation occurred when SNP were used compared with using pedigree. For example, in the small data set, the standard deviation decreased from 0.22 to 0.18 for the correlation between BW and milk. Therefore, it can be concluded that using genomic relationships improved the accuracy of the estimated genetic correlations.

DISCUSSION

The objective of this study was to estimate variance components for milk yield, DMI, and BW in a small data set using genetic relationships based on SNP or pedigree information, and compare the precision of the estimated genetic parameters. For this purpose, standard errors from ASReml were compared with the sampling standard deviation from a parametric bootstrap procedure developed in this study. The estimated heritabilities reported here are generally higher than those found elsewhere (Veerkamp, 1998), especially those estimated with the pedigree relationships. Tight management and relatively homogeneous groups of animals might be an explanation for a low residual variance and, therefore, a higher heritability. In addition, the typical structure with full siblings and daughter-dam relationships might have inflated the heritability, as environmental covariances might have existed between relatives, and these were not modeled explicitly

Table 4. Mean of the 250 replicates for the heritability, genetic covariances (below diagonal), variances (underlined on diagonal), and genetic correlations (above diagonal), using numerator (**A517** and **A639**), genomic (**G517**), and combined (**H639**) relationship matrices when the same model was used in simulation and analysis¹

	A517			G517		
	Milk	DMI	BW	Milk	DMI	BW
Milk (kg/d)	<u>7.3</u>	0.36	0.10	<u>7.5</u>	0.35	0.09
DMI (kg/d)	1.3	<u>1.8</u>	0.70	1.3	<u>1.8</u>	0.70
BW (kg)	7.0	25.7	<u>746</u>	5.9	25.0	<u>727</u>
h^2	0.42	0.65	0.50	0.43	0.64	0.49
	A639			H639		
	Milk	DMI	BW	Milk	DMI	BW
Milk (kg/d)	<u>7.3</u>	0.35	0.11	<u>7.5</u>	0.36	0.09
DMI (kg/d)	1.3	<u>1.8</u>	0.71	1.3	<u>1.8</u>	0.71
BW (kg)	7.5	25.9	<u>744</u>	6.7	26.1	<u>752</u>
h^2	0.41	0.65	0.50	0.43	0.65	0.51

¹Simulated parameters are in Table 1.

Table 5. Mean and standard deviation of the 250 replicates for the heritability, genetic covariances (below diagonal), variances (underlined on diagonal), and genetic correlations (above diagonal), using the numerator (**A517**) relationship matrix when different models were used in simulation and analysis¹

	Simulation: A517 ; analysis: G517			Simulation: G517 ; analysis: A517		
	Milk	DMI	BW	Milk	DMI	BW
Mean						
Milk (kg/d)	<u>4.9</u>	0.34	0.09	<u>7.8</u>	0.37	0.11
DMI (kg/d)	0.8	<u>1.2</u>	0.69	1.4	<u>1.9</u>	0.70
BW (kg)	4.2	17.2	<u>508</u>	8.0	26.8	<u>775</u>
h^2	0.29	0.45	0.35	0.43	0.66	0.51
SD						
Milk (kg/d)	<u>1.6</u>	0.21	0.24	<u>2.3</u>	0.17	0.23
DMI (kg/d)	0.5	<u>0.3</u>	0.14	0.8	<u>0.4</u>	0.12
BW (kg)	11.2	6.1	<u>165</u>	17.3	7.6	<u>206</u>
h^2	0.09	0.10	0.10	0.11	0.11	0.12

¹Simulated parameters are in Table 1.

(Meyer, 1992). Also, some of the animals were part of the CRV breeding program and in the analysis here, no adjustments were made for this genetic group effect as was previously done by Veerkamp et al. (2000). All of these factors might have inflated the estimates of all heritabilities in the present study.

When using the genomic relationships, heritability estimates decreased. Ideally, this is because some of the environmental effects confounded with the pedigree relationships are not included in the genetic variances anymore, as was observed in mice data (Lee et al., 2010). However, several reasons exist why estimated genetic variances differ between the models using pedigree and genomic relationships. First of all, the 2 relationship matrices are different. For example, when comparing the diagonal elements of the 2 matrices for the 517 animals (Figure 2), a clear difference in scale was observed. Also, the sum of the eigenvalues for **A517** and **G517** was 537 and 512, respectively. Simply rescaling matrix **G517** to **A517** based on the eigenvalues moved the genetic variances marginally and did not bring the results from **A517** and **G517** closer together [e.g., genetic variances moved from 7.8 to 7.5 for **G517**, where **A517** gave 9.3 (full results not shown)]. This is because fundamental differences exist in the genetic models underlying the ARM and GRM matrices (Powell et al., 2010). The relationships in the pedigree matrix are based on identity by descent (i.e., the probability that 2 alleles come from a common ancestor). The relationships in the genomic matrix reflect the identity by state probabilities (i.e., the probability that 2 alleles are the same). Powell et al. (2010) demonstrated that variance components estimated using the genomic matrix are for an arbitrary and undefined base population. This is in contrast with a model using the pedigree relationships, where the

variance components are estimated in the defined base population of the pedigree. These different bases and scales of the ARM and GRM are, together with the information on the Mendelian sampling component in the GRM, possible explanations for the different variance components. This also poses the question what is the most relevant base for combining the ARM and GRM in an **H** matrix. Variance components estimated with **H** (i.e., the combined **G** and **A**), are therefore, also for an arbitrary base population. Although, in the present study, results look promising and it appears worthwhile combining the data of genotyped and non-genotyped animals, more proper scaling methods are required for combining **G** and **A** to estimate variance components; for example, a rescaling based on observed allele frequencies was suggested by Forni et al. (2011).

Using SNP-based relationships improved the precision of the heritability estimate clearly. The bootstrap information suggested that this was not because of the different level of the estimated heritability, as the different models gave the same values for the heritability. The increase in the precision was such that, in our situation, it was more advantageous to genotype 517 animals, rather than add 122 new phenotypes. Therefore, in some scenarios it might be worthwhile genotyping animals, rather than collecting more data. As expected, genotyping the animals and collecting additional data (**H639**) gave the most precise estimates. Surprisingly, no advantage for the estimation of genetic correlations was observed when using the SNP-based relationships compared with the pedigree relationships when using the standard errors from ASReml. Generally speaking, experiments that improve the estimation of the heritability also improve the estimation of the genetic correlation (Falconer and Mackay, 1996) because the standard error of the genetic correlation is approximated by

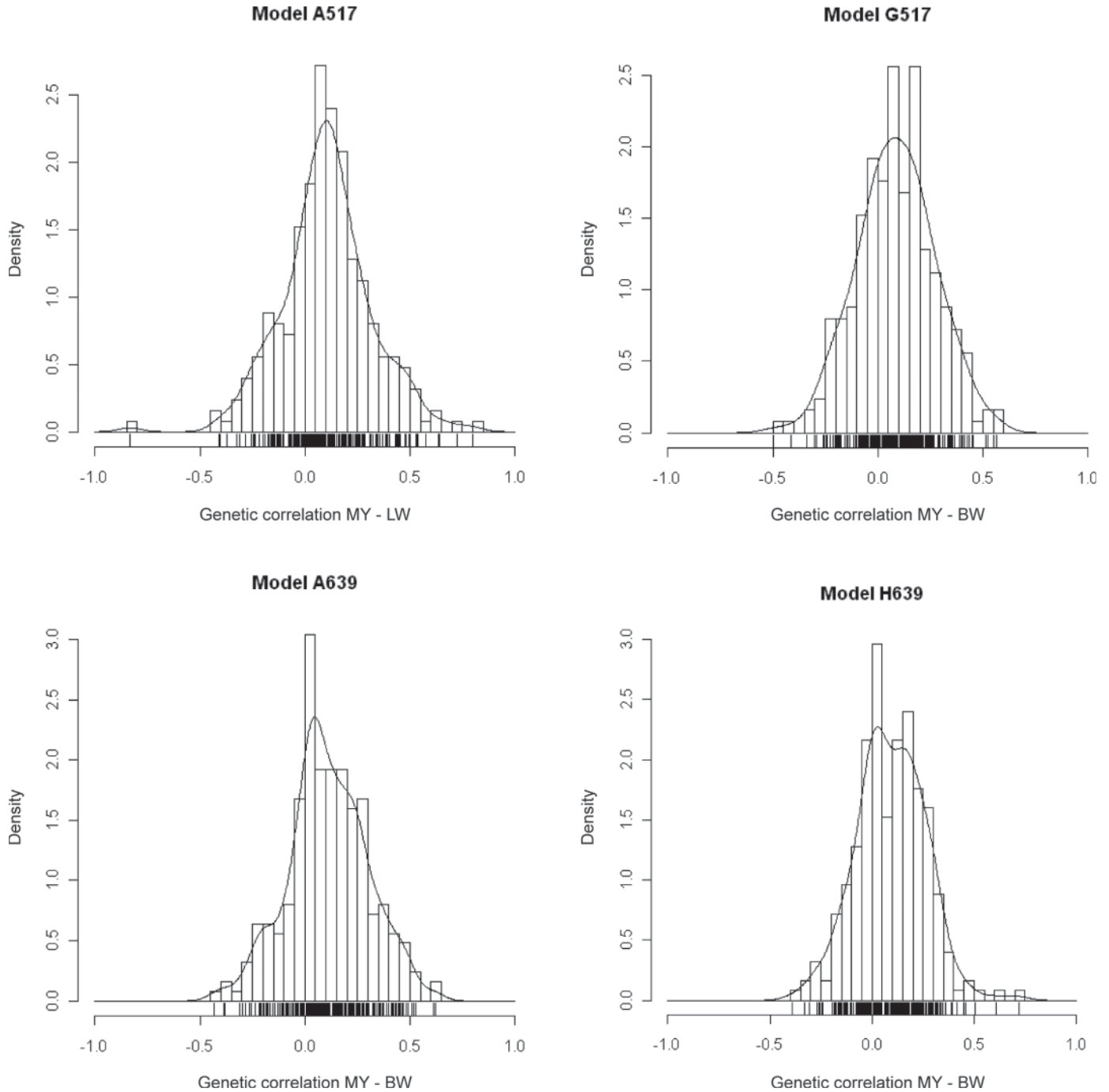


Figure 1. Sampling distribution of the genetic correlation between milk yield (MY) and BW across the 250 replicates, using numerator (**A517** and **A639**), genomic (**G517**), and combined (**H639**) relationship matrix.

$$\frac{1 - r_g^2}{\sqrt{2}} \sqrt{\frac{se_{h1} se_{h2}}{h_1^2 h_2^2}}$$

where r_g is the genetic correlation between traits 1 and 2, and h_1^2 and h_2^2 are the heritability for traits 1 and 2, with the respective standard errors se_{h1} and se_{h2} . This

formula also shows that for genetic correlations close to unity, a small decrease in the estimate causes a large increase in the standard error. Using the estimates of model **A517** and this equation to predict the expected value for the standard error of the genetic correlation between milk and BW, a value of 0.139 was predicted. When the values of model **G517** were used, the pre-

Table 6. Standard deviation of the 250 replicates for the heritability, genetic covariances (below diagonal), variances (underlined on diagonal), and genetic correlations (above diagonal), using numerator (**A517** and **A639**), genomic (**G517**), and combined (**H639**) relationship matrices

	A517			G517		
	Milk	DMI	BW	Milk	DMI	BW
Milk	<u>2.2</u>	0.17	0.22	<u>2.1</u>	0.15	0.18
DMI	0.7	<u>0.4</u>	0.13	0.7	<u>0.4</u>	0.09
BW	14.7	7.9	<u>220</u>	12.9	6.2	<u>173</u>
h^2	0.11	0.12	0.13	0.10	0.09	0.10
	A639			H639		
	Milk	DMI	BW	Milk	DMI	BW
Milk	<u>2.1</u>	0.15	0.19	<u>1.8</u>	0.13	0.17
DMI	0.6	<u>0.4</u>	0.09	0.6	<u>0.3</u>	0.09
BW	13.5	6.1	<u>174</u>	12.0	5.9	<u>160</u>
h^2	0.10	0.09	0.10	0.09	0.09	0.09

dicted standard error was 0.145, implying an increase of 0.006. For the genetic correlation between milk and DMI, and between DMI and BW, increases in the standard error were predicted to be from 0.003 and 0.005 when replacing pedigree relationships by genomic relationships. This general effect of the level of the genetic correlation on its standard error explains why, in our results, an increase in the standard error occurred when using the GRM. This is because the scale differences resulted in a lower level of the genetic correlations (i.e., closer to zero) estimated with the GRM compared with those estimated with the ARM. This also explains why the bootstrapping standard deviation did improve with

SNP-based relationships. Therefore, when compared at the same level of the genetic correlations, the GRM gave also more accurate estimates of the genetic correlations than the ARM.

The parametric bootstrapping procedure was useful in this small data set to investigate the sampling properties of the models, and results were close to the estimates of the standard error of ASReMl. Because the data was simulated using equivalent relationship matrices and the same variance components, the estimated sampling variance was purely due to information content of the model fitted. Differences between the traits that might exist in the underlying genetic

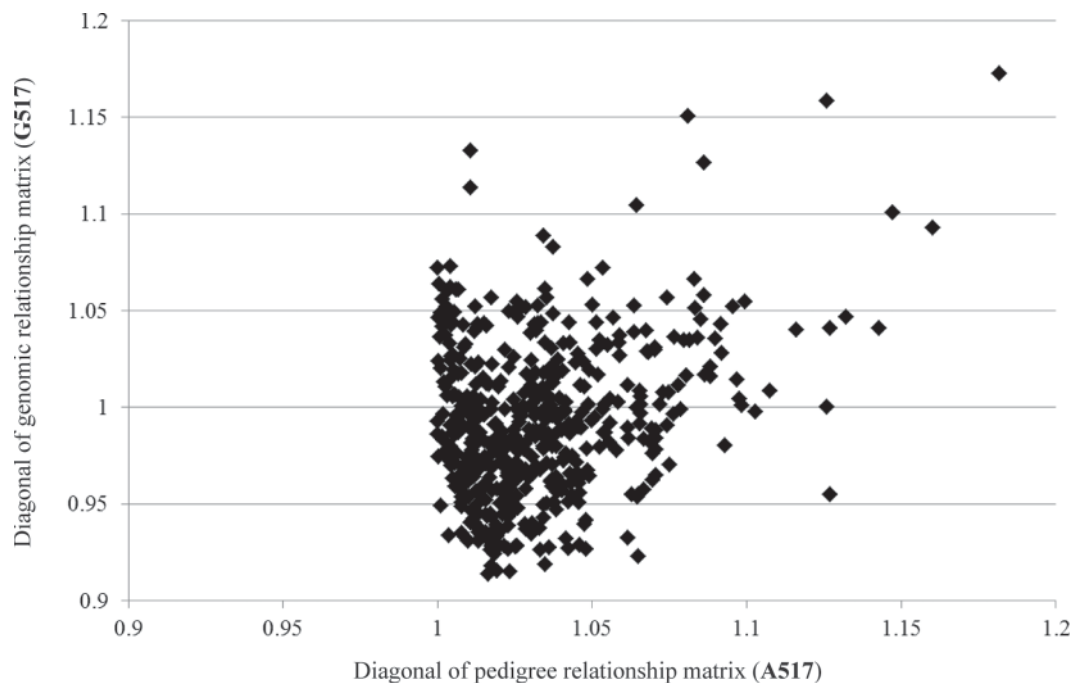


Figure 2. Comparison of the diagonal elements of the genomic relationships matrix (**G517**) with the diagonal elements of the same animals in the pedigree relationship matrix (**A517**).

model, or how each model fits the data are ignored in the bootstrapping procedure. For example, for some traits, single genes with a large effect might be present, but in all 4 models used here, an underlying genetic model of many genes with a small effect on each trait was assumed. Hence, SNP information was not used to identify QTL with a large effect, or to take into account different genes affecting each trait. Hence, it is a typical infinitesimal model approach, with equal weight to each SNP in determining the relationships between animals. One step further could be to expand the model with genomic relationships to a model where a different weight is given to each SNP (VanRaden, 2008; Goddard, 2009; Veerkamp et al., 2010). The genomic relationship can then be weighted based on the estimated effect of each SNP on a trait.

CONCLUSIONS

Using genomic relationships improved the precision of the estimated genetic parameters (heritability and genetic correlations between traits). In the present study, it was more advantageous to genotype 517 animals, rather than use 122 more phenotypes. Animals without genotype could be included in the analysis by merging genomic and pedigree relationships. The parametric bootstrapping procedure was a useful tool in the small data set to investigate the sampling properties.

ACKNOWLEDGMENTS

This project is partly financed by the RobustMilk project, which is financially supported by the European Commission under the Seventh Research Framework Program, Grant Agreement KBBE-211708.

REFERENCES

Aguilar, I., I. Misztal, D. L. Johnson, A. Legarra, S. Tsuruta, and T. J. Lawlor. 2010. Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *J. Dairy Sci.* 93:743–752.

Calus, M. P. L., T. H. E. Meuwissen, A. P. W. de Roos, and R. F. Veerkamp. 2008. Accuracy of genomic selection using different methods to define haplotypes. *Genetics* 178:553–561.

Falconer, D. S., and T. F. C. Mackay. 1996. *Introduction to Quantitative Genetics*. 4th ed. Longman Group, Essex, UK.

Forni, S., I. Aguilar, and I. Misztal. 2011. Different genomic relationship matrices for single-step analysis using phenotypic, pedigree and genomic information. *Genet. Sel. Evol.* 43:1–8.

Genz, A., F. Bretz, T. Miwa, X. Mi, F. Leisch, F. Scheipl, B. Bornkamp, and T. Hothorn. 2010. *Multivariate normal and t distributions. Computes multivariate normal and t probabilities, quantiles, random deviates and densities*. Accessed Sept. 12, 2010. <http://cran.r-project.org/web/packages/mvtnorm/mvtnorm.pdf>.

Gilmour, A. R., B. R. Cullis, S. J. Welham, and R. Thompson. 2000. *ASReml. Program User Manual*. NSW Agriculture, Orange Agricultural Institute, Orange, NSW, Australia.

Goddard, M. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. *Genetica* 136:245–257.

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, and M. E. Goddard. 2009. Invited review: Genomic selection in dairy cattle: Progress and challenges. *J. Dairy Sci.* 92:433–443.

Lee, S. H., M. E. Goddard, P. M. Visscher, and J. H. J. van der Werf. 2010. Using the realized relationship matrix to disentangle confounding factors for the estimation of genetic variance components of complex traits. *Genet. Sel. Evol.* 42:22–36.

Legarra, A., and I. Misztal. 2008. Computing strategies in genome-wide selection. *J. Dairy Sci.* 91:360–366.

Meuwissen, T. H. E., and Z. Luo. 1992. Computing inbreeding coefficients in large populations. *Genet. Sel. Evol.* 24:305–313.

Meyer, K. 1992. Bias and sampling covariances of estimates of variance components due to maternal effects. *Genet. Sel. Evol.* 24:487–509.

Powell, J. E., P. M. Visscher, and M. E. Goddard. 2010. Reconciling the analysis of IBD and IBS in complex trait studies. *Nat. Rev. Genet.* 11:800–805.

VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.

Veerkamp, R. F. 1998. Selection for economic efficiency of dairy cattle using information on live weight and feed intake: A review. *J. Dairy Sci.* 81:1109–1119.

Veerkamp, R. F., H. A. Mulder, and M. P. L. Calus. 2010. Estimation of heritability for dairy traits, combining pedigree with dense SNP information on some animals. *Proc. 9th World Conference on Genetics Applied to Livestock Production Leipzig, Germany*.

Veerkamp, R. F., J. K. Oldenbroek, H. J. Van Der Gaast, and J. H. J. Van Der Werf. 2000. Genetic correlation between days until start of luteal activity and milk yield, energy balance and live weights. *J. Dairy Sci.* 83:577–583.

Verbyla, K. L., M. P. L. Calus, H. A. Mulder, Y. de Haas, and R. F. Veerkamp. 2010. Predicting energy balance for dairy cows using high-density single nucleotide polymorphism information. *J. Dairy Sci.* 93:2757–2764.

Yang, J., B. Benyamin, B. P. McEvoy, S. Gordon, A. K. Henders, D. R. Nyholt, P. A. Madden, A. C. Heath, N. G. Martin, G. W. Montgomery, M. E. Goddard, and P. M. Visscher. 2010. Common SNPs explain a large proportion of the heritability for human height. *Nat. Genet.* 42:565–569.