The estimation of genomic relationships using breedwise allele frequencies among animals in multibreed populations

M. L. Makgahlela,*†¹ I. Strandén,*† U. S. Nielsen,‡ M. J. Sillanpää,§ and E. A. Mäntysaari†

*Department of Agricultural Sciences, PO Box 27, FIN-00014, University of Helsinki, Helsinki, Finland

†MTT Agrifood Research Finland, Biotechnology and Food Research, Biometrical Genetics, FIN-31600 Jokioinen, Finland

‡Danish Agricultural Advisory Service, Udkaersvej 15, 8200 Aarhus, Denmark

§Department of Mathematical Sciences and Department of Biology and Biocenter Oulu, PO Box 3000, FIN-90014, University of Oulu, Oulu, Finland

ABSTRACT

Different approaches of calculating genomic measures of relationship were explored and compared with pedigree relationships (**A**) within and across base breeds in a crossbreed population, using genotypes for 38,194 loci of 4,106 Nordic Red dairy cattle. Four genomic relationship matrices (**G**) were calculated using either observed allele frequencies (AF) across breeds or within-breed AF. The **G** matrices were compared separately when the AF were estimated in the observed and in the base population. Breedwise AF in the current and base population were estimated using linear regression models of individual genotypes on breed composition. Different **G** matrices were further used to predict direct estimated genomic values using a genomic BLUP model. Higher variability existed in the diagonal elements of **G** across breeds (standard deviation = 0.06, on average) compared with **A** (0.01). The use of simple observed AF across base breeds to compute **G** increased coefficients for individuals in distantly related populations. Estimated breedwise AF reduced differences in coefficients similarly within and across populations. The variability of the current adjusted **G** matrix decreased from 0.055 to 0.035 when breedwise AF were estimated from the base breed population. The direct estimated genomic values and their validation reliabilities were, however, unaffected by AF used to compute **G** when estimated with a genomic BLUP model, due to inclusion of breed means in the model. In multibreed populations, **G** adjusted with breedwise AF from the founder population may provide more consistency among relationship coefficients between genotyped and ungenotyped individuals in an across-breed single-step evaluation.

Key words: allele frequency, genomic relationship matrix, admixed population, genomic prediction

INTRODUCTION

The use of marker genotypes to estimate relationships among individuals in a population has become increasingly important in many fields of genetics. In livestock breeding, knowledge of relationships is used routinely to estimate genetic variation and animal breeding values (EBV; VanRaden, 2008; Hayes et al., 2009; Su et al., 2012), monitor inbreeding (Fernández et al., 2011; Toro et al., 2011), and for conservation of animal genetic resources (Eding and Meuwissen, 2001). Traditionally, relationship coefficients are calculated from the pedigree data. Pedigree relationships are obtained as 2 times the expected average identity by descent (**IBD**) sharing between 2 relatives (Malécot, 1948) and have been applied successfully within the framework of mixed-model equations for best linear unbiased prediction of EBV. Presently, with the increasing availability of genetic markers covering the whole genome, pedigree-based relationships can be replaced or combined with realized relationships calculated from marker data in the prediction of genomic breeding values (genomic EBV; Habier et al., 2007; Hayes et al., 2009).

Realized relationships derived from molecular markers are based on the actual **IBD** sharing or identity by state for genomic regions and, therefore, have more variation between closely related animals than pedigree relationships (VanRaden, 2008; Hayes et al., 2009). Moreover, realized relationships capture unrecorded pedigrees. Several different methods of calculating genomic relationship matrices (**G**) have been developed, for genotyped animals only (VanRaden, 2008; Yang et al., 2010) and when genotyped and ungenotyped individuals are combined (Legarra et al., 2009; Misztal et al., 2009; Christensen and Lund, 2010). In the latter approach, an arbitrary weight on pedigree relationships (**A**) is often used to measure the amount of variation not explained by markers. Although variability exists in the accuracy of predictions among the above methods (Forni et al., 2011), generally these accuracies are at

Received December 24, 2012.

Accepted April 24, 2013.

¹Corresponding author: mahlako.makgahlela@helsinki.fi

least twice of those estimated from pedigree data only (see, for example, Su et al., 2012).

Genomic relationship matrix **G** can be constructed by using a matrix having genotype information for each individual and marker (VanRaden, 2008). Each genotype is a deviation from the marker-specific population mean, which is calculated using population allele frequencies. The estimation of **G** has been shown to be closer to **A** when inferences use allele frequencies (**AF**) in the distant ancestral population (VanRaden, 2008; VanRaden et al., 2011) instead of **AF** in the currently genotyped population. This is because the expected **G** coefficients would be expressed relative to the same base population as **A**. The limitation is that base population **AF** are generally not available with field data and their estimation can be challenging. Gengler et al. (2007) showed an efficient way of calculating gene content and base population **AF** within a breed. However, in practice, currently genotyped populations are assumed to be the base population. This results from centering the genotype matrix used to build **G** with current-data **AF** so that the average genomic relationship between animals within the current population become 0 and scaling **G** such that the additive genetic variance would be comparable to that obtained through conventional methods (Powell et al., 2010; Forni et al., 2011).

The use of observed **AF** within a breed may not have major practical implications in genomic BLUP (**GBLUP**) models. In the context of structured populations, the effect of using across-breed **AF** to make **G** may have consequences for the estimation of relationships, mainly attributable to varying source of **AF** between breeds. Eding and Meuwissen (2001) demonstrated that average relatedness between 2 populations could be expressed in terms of population-specific **AF**. VanRaden et al. (2011) used the average of 3 breedwise **AF** for estimation in the combined 3-breed population. Although these approaches would be beneficial for multiple populations with distinctive subpopulations, a need still exists for approaches in populations that constitute mainly crossbred animals. The Nordic Red dairy cattle (**RDC**) comprise 3 subpopulations by country of birth [i.e., Denmark (**DNK**), Sweden (**SWE**), and Finland (**FIN**)]. Over the years of crossbreeding, the majority of animals (~98%) in the Nordic RDC are composites of base breeds. The absence of pure base-breed animals remains a major challenge for the estimation of breedwise **AF**. The objective of this study was to investigate whether the use of estimated breedwise **AF** in the calculation of genomic relationships would provide a more accurate estimate of **G** than using **AF** across breeds, and to determine the effect on **G** when **AF** are estimated in the base population versus the currently genotyped population.

MATERIALS AND METHODS

Data

This study was carried out in a structured population with 60 pure base breed bulls and 4,046 bulls of combinations of base breeds in the Nordic RDC. Genotypes for all 4,106 bulls were attained using the Illumina Bovine SNP50 BeadChip (Illumina Inc., San Diego, CA). For quality purposes, markers from the X chromosome, without map position in the UMD3.0 genome assembly (Zimin et al., 2009) and with minor allele frequency (**MAF**) <5% were discarded. In addition, animal genotypes with a GenCall score (Illumina Inc., 2005) <60% and marker loci with call rates <5% in a large reference sample from the same genotyping laboratory, consisting of Danish Holstein bulls, were discarded. Finally, missing genotypes were imputed using fastPHASE software chromosome by chromosome (Scheet and Stephens, 2006). Due to unavailability of pure base breeds, informative SNP above were selected based on across-breed **AF**. After quality control, a total of 38,194 SNP markers were available for analyses. The entire RDC pedigree, containing over 4 million records, was used to calculate breed proportions (**BP**) for individual bulls (Lidauer et al., 2006). A breed was defined only if the average **BP** in the data was greater than 10%. Breeds used in this study were the Swedish Red (**SRB**), Finnish Ayrshire (**FAY**), and Norwegian Red (**NRF**), and the remaining breeds with **BP** less than 10% were combined into the breed "other." A more detailed description about the population structure, breeds contained, and definitions of the final 4 breeds and their trends is provided by Makgahlela et al. (2013). The pedigree for genotyped bulls contained 22,300 animals.

Phenotypes were individual daughter deviations (**IDD**) of 1,995,606 RDC cows for milk, protein, and fat yields, obtained from March 2010 official evaluations of the Nordic cattle genetic evaluations. By definition, the **IDD** are cow performances adjusted for fixed effects, nongenetic random effects, and genetic effects of the cow's dam (Mrode and Swanson, 2004). Here, however, **IDD** were computed by using animal model deregression from 305-d combined EBV (Mäntysaari et al., 2011). For validation of predictions using different **G** matrices, the data were split into sets of 3,300 training bulls born between 1980 and 1999, and 806 validation bulls born between 1998 and 2005. The training data had older bulls, which were evaluated for the first time during the 2005 Nordic cattle routine evaluation.

Estimation of Pedigree and Genomic Relationships

Pedigree relationships of genotyped bulls were estimated using the RelaX2 computer program (Strandén

and Vuori, 2006). Genomic relationships were computed following methods 1 and 2 as demonstrated by VanRaden (2008) and modifications of the 2 methods to adapt to the admixed structure of the current population.

Let there be n animals that have been genotyped for m markers. Let u_{ij} be genotype j of animal i , where genotype u_{ij} is the number of copies for the second allele (i.e., u_{ij} has value 0, 1, or 2). Following VanRaden (2008) method 1 and using observed AF across breeds, the original genomic relationship matrix $\mathbf{G}$ (denoted **Gorg**) was computed as $\mathbf{Gorg} = \mathbf{ZZ}'/k$, where $\mathbf{Z}$ is an n by m matrix of centered genotypes. Here, the element of animal i for marker j in $\mathbf{Z}$ is $u_{ij} - 2p_j$, where p_j is the frequency of the second allele at SNP marker j and $k = 2\sum_j p_j(1 - p_j)$. In method 2, also shown by Yang et al. (2010), standardized genotypes were used to calculate the $\mathbf{G}$ matrix as $\mathbf{Gorg2} = \mathbf{Z}^*\mathbf{Z}'^*/m$, where each column j in $\mathbf{Z}^*$ is $\mathbf{Z}_j^* = \mathbf{Z}_j / \sqrt{2p_j(1 - p_j)}$ and $\mathbf{Z}_j$ is column j in $\mathbf{Z}$.

Two adjusted genomic matrices were computed using breedwise AF: **Gadj** and **Gadj2**. These matrices were obtained by modifying methods 1 and 2 by VanRaden (2008). Matrix **Gadj** was calculated as $\mathbf{Gadj} = \mathbf{MM}'/k$, where, with the same notation as in $\mathbf{Z}$, elements of $\mathbf{M}$ are $u_{ij} - 2p_{ij}$ and k is assumed the same as above. Here, the p_{ij} is the allele frequency for the j th SNP marker, expected for genotype M_{ij} and taking into account the breed background of animal i . For the matrix **Gadj2**, the columns of $\mathbf{M}$ were further scaled by the standard deviation of the expected marker effects to obtain the $\mathbf{M}^*$ matrix. For animal i , the genotype element j of $\mathbf{M}^*$ was $\frac{u_{ij} - 2p_{ij}}{\sqrt{2p_{ij}(1 - p_{ij})}}$. To improve numerical stability and avoid division by 0, estimated breedwise AF below 0.05 or above 0.95 were set to either 0.05 or 0.95, respectively. This threshold corresponded to our prior removal of SNP with MAF <5%, which was based on across-breed AF. Then finally, the relationship matrix was obtained as $\mathbf{Gadj2} = \mathbf{M}^*\mathbf{M}^*/m$. Individual AF p_{ij} was calculated using the currently genotyped animals or base population as reference population.

Current genotyped population-level breedwise AF were computed using a linear multiple regression and binomial models. A simple multiple regression vector (β) of genotypes ($\mathbf{y}$) on breed proportions ($\mathbf{X}$) was solved for each marker, and $\mathbf{e}$ is the independently normally distributed residual error:

$$\mathbf{y} = \mathbf{X}\beta + \mathbf{e}.$$

Alternatively, genotype y_i for individual i was considered as observation from binomial (p_{ij} , 2). The binomial likelihood $\prod_{i=1}^n p_{ij}^{y_{i1}} (1 - p_{ij})^{y_{i2}}$ was handled by having a logistic regression for parameters p_{ij} , defined as follows:

$$\text{logit}([p_{ij}]) = \mathbf{X}\beta,$$

where $\text{logit}(p_{ij}) = \ln\left(\frac{p_{ij}}{1 - p_{ij}}\right)$. The expected AF of the marker for each individual from the linear and logistic models become $[\hat{p}_{ij}] = \mathbf{X}\hat{\beta}$ and $[\hat{p}_{ij}] = \exp \mathbf{X}\hat{\beta} / [1 + \exp(\mathbf{X}\hat{\beta})]$, respectively, where AF in $\hat{\beta} = (\hat{\beta}_1, \dots, \hat{\beta}_4)$ in for SRB, FAY, NRF, and the combined breeds "other."

The AF across breeds and breedwise AF were estimated in the base (founding) population using the gene content approximation algorithm of Gengler et al. (2007). The setup in this algorithm follows the logic of genetic covariance among relatives, where the covariance between gene contents, which is the number of copies of one allele in a genotype, is proportional to the additive relationship between animals. Large pedigrees are used to compute $\mathbf{A}$ and linear mixed-model equations are used to account for selection and drift in AF across time, thus occurring during pedigree generations (Gengler et al., 2007). As was done for the simple linear model above, the expected SNP genotype for ungenotyped base population animals was estimated from their genotyped relatives for every marker following the following model:

$$\mathbf{y} = \mathbf{X}\gamma + \mathbf{Q}\mathbf{g} + \mathbf{e},$$

where γ is the vector of breed effects or intercept and $\mathbf{Q}$ is a design matrix allocating records to animal effects $\mathbf{g}$, which is the estimated gene content for all animals including ancestors. The regression effect solutions to γ were used to calculate individual AF values p_{ij} similarly as using the current genotyped population AF.

The constructions of all 4 $\mathbf{G}$ above (i.e., **Gorg**, **Gorg2**, **Gadj**, and **Gadj2**) were computed using corresponding AF estimated in the currently genotyped population. Additionally, **Gorg**, **Gadj**, and **Gadj2** relationship matrices were recalculated using corresponding AF estimated from the base population. For the comparison in the calculation of within-breed AF, genotypes of the oldest 842 bulls in the data, born between 1971 and 1990, were used to estimate base population breedwise AF. Here, we used a simple mul-

multiple regression of genotypes on breed proportions as explained for the currently genotyped population.

Effect of Alternative G Estimates on Genomic Predictions

The estimation of variance components and the prediction of direct estimated genomic values (**DGV**) were carried out separately for each of the 3 matrices (i.e., **Gorg**, **Gadj**, and **Gadj2**) and separately when they were calculated using either observed or base population AF. The analyses were conducted using ASReml 3.0 (Gilmour et al., 2009) and MiX99 (Lidauer and Strandén, 1999) software under the following GBLUP model:

$$\mathbf{IDD} = \mathbf{Xb} + \mathbf{Sa} + \mathbf{e},$$

where **IDD** is a vector of IDD for daughters of bulls in the reference data set, **b** is a vector of breed effects or intercept, **S** is the design matrix that relates observations to DGV for sires of daughters **a**, and **e** is a vector of random normal deviates. It is assumed that $\mathbf{e} \sim N(0, \mathbf{R}\sigma_e^2)$, where the diagonal element in **R** is $r_{ii} = 1/w_i$, and w_i is the number of effective record contributions of the cow and is a weighting factor for the i th IDD. It is assumed that $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$, where **G** is the marker-based relationship matrix and σ_a^2 is the additive genetic variance. Here the fixed-breed regression effects were only fitted for **Gadj** and **Gadj2**. The predicted values for all animals in this case were obtained as the sum of the animals' DGV and fixed-breed regression solutions.

Validation Analyses

The validation reliability of DGV was assessed following the Interbull genomic EBV validation test (Mäntysaari et al., 2010). A weighted regression model of deregressed breeding value (**DRP**) for bulls in the validation data on predicted DGV was fitted to obtain the regression b_1 coefficient:

$$\mathbf{DRP} = 1b_0 + b_1\hat{\mathbf{a}} + \mathbf{e},$$

where **DRP** is the vector of **DRP** for the candidate bulls and $\hat{\mathbf{a}}$ is the vector of estimated DGV for these bulls. The linear model was weighted by individual w_k , defined as the effective daughter contribution of the bull. The validation reliability of the model was calculated as $R_{\text{DGV}}^2 = \frac{r_{(\text{DRP}, \text{DGV})}^2}{\bar{w}}$, where $r_{(\text{DRP}, \text{DGV})}^2$ is the

squared correlation between **DRP** and **DGV** and $\bar{w}$ is the average of w_k , which account for the inaccuracy in the estimation of **DRP**.

RESULTS

Estimated Breedwise AF

Breedwise AF were calculated for the defined breeds SRB, FAY, NRF, and the breed "other," which combined small breeds. Only 60 bulls were pure base breed, with 59 having 100% BP for FAY. Few individuals had a BP of at least 50% for SRB (647) and NRF (40) in the data. In the genotyped population, both the linear regression and binomial models gave equivalent estimates of AF for the breeds, with correlations over all SNP between models close to 1. Whereas a linear model resulted in AF for few markers outside the expected range of 0 and 1, a binomial model restrained coefficients to fall within this range. The distributions of breedwise AF were also generally similar under the linear versus binomial model, except the NRF having the most markers with estimated AF out of the expected range using the linear model. A binomial model was, however, challenging to implement using Gengler's method due to software limitations. Because the estimated AF were similar, for consistency we present results from a linear regression model. Table 1 shows the numbers of markers for each breed with estimated AF that were estimated to be either less than 0 or greater than 1 when using the linear model. The least number of markers outside the parameter space were found when AF were estimated in the base population, with the range of 313 below 0 and 1,972 above 1. The most markers with AF out of parameter space were found when 842 old bulls were used in the AF estimation. Then, 2,460 AF were below 0 and 8,387 were greater than 1. The breeds NRF and SRB appeared to have the most markers with estimated AF outside the parameter space, which could partly be due to not having 100% NRF cattle and only 1 pure base breed SRB cattle in the genotyped population.

The between-breed correlations of breedwise AF estimated in the base population, currently genotyped population, and the oldest genotyped bulls are shown in Table 2. Generally, breedwise AF had the highest correlations when AF were estimated in the base population (ranging from 0.678 to 0.817) and least correlated when estimated using genotypes from older bulls (0.421 to 0.651). The highest correlations in the base population AF estimates were between SRB and NRF (0.817), and lowest correlations were between NRF and the breed "other" (0.678). However, observed AF estimated using all genotypes and old genotyped bulls were highly

Table 1. The numbers of markers with estimated allele frequencies that fell outside the expected range of 0 and 1 for each breed when estimated with breed proportions as genetic group in the gene content model (base population) or in the linear regression model (current population)

Item	Breed ¹			
	SRB	FAY	NRF	Other
Base population				
Less than 0	26	19	265	3
Greater than 1	276	193	1,449	54
Currently genotyped population				
Less than 0	152	94	960	7
Greater than 1	1,615	902	3,304	81
Genotypes from 842 old bulls				
Less than 0	275	78	2,089	18
Greater than 1	1,848	668	5,708	163

¹SRB = Swedish Red; FAY = Finnish Ayrshire; NRF = Norwegian Red; other = combined breeds.

correlated between SRB and FAY at 0.643 and 0.630, respectively, and less correlated between SRB and NRF (0.545 and 0.421, respectively).

Estimated Relationship Coefficients from Pedigree and Genomic Data

Descriptive statistics of diagonal elements from the pedigree ($\mathbf{A}_{ii}$) and different genomic estimators ($\mathbf{G}_{ii}$) are presented in Table 3, across populations and within bulls registered in DNK, SWE, and FIN. The number of bulls was 800, 1,240, and 2,040 in the DNK, SWE, and FIN populations, respectively. Results are presented by country because these 3 populations are generally considered a single population; however, the genetic relationships between SWE and FIN populations is stronger than these 2 and the DNK population. The ranges of maximum elements from $\mathbf{A}$ within (1.081 to 1.135) and across populations (1.135) were smaller

than those from $\mathbf{G}$ estimators within (1.233 to 1.450) and across populations (1.310 to 1.450). This difference in scales is because $\mathbf{A}$ is not an absolute measurement but an expected relatedness given the pedigree, whereas $\mathbf{G}$ measures the actual relatedness at marker loci. The variability of $\mathbf{G}$ across populations was greater in the original approaches (i.e., **Gorg** and **Gorg2**) and smaller for adjusted matrices, especially for **Gadj2**. Similar tendencies were found within the DNK animals but not for SWE and FIN bulls. Coefficients from **Gorg** and **Gorg2** were generally similar, but we present estimates from both methods 1 and 2 for comparison to their adjusted alternatives proposed in the current study. The average $\mathbf{A}_{ii}$ was greatest in the FIN bulls (1.016) and smallest in the DNK bulls (1.007). However, these averages were vice versa for $\mathbf{G}_{ii}$ using observed AF from **Gorg**. The mean of diagonals from $\mathbf{A}$ and **Gorg** were close to 1 for across breeds in SWE and FIN but was 1.136 for DNK from **Gorg**.

Table 2. The correlations between breedwise allele frequencies estimated in the base population (Gengler et al., 2007) within the oldest 842 bulls and in the currently genotyped population

Item	Breed ¹			
	SRB	FAY	NRF	Other
Base population				
SRB	1.000	0.737	0.817	0.741
FAY		1.000	0.688	0.771
NRF			1.000	0.678
Other				1.000
Currently genotyped population				
SRB	1.000	0.644	0.602	0.626
FAY		1.000	0.545	0.672
NRF			1.000	0.570
Other				1.000
Genotypes from old bulls				
SRB	1.000	0.630	0.468	0.594
FAY		1.000	0.421	0.651
NRF			1.000	0.446

¹SRB = Swedish Red; FAY = Finnish Ayrshire; NRF = Norwegian Red; other = combined breeds.

Table 3. Descriptive statistics of diagonal elements from the pedigree (**A**), original, and adjusted genomic relationship matrices (**G**) estimated using allele frequencies in the genotyped population¹

Item	Mean	Minimum	Maximum	SD
Across populations				
A	1.012	1.000	1.135	0.014
Gorg ²	1.019	0.871	1.379	0.074
Gorg 2 ²	1.019	0.871	1.379	0.074
Gadj ³	0.949	0.747	1.310	0.045
Gadj 2 ³	0.965	0.773	1.450	0.055
Danish bulls				
A	1.007	1.000	1.109	0.013
Gorg ²	1.136	0.973	1.328	0.072
Gorg 2 ²	1.136	0.973	1.328	0.072
Gadj ³	0.960	0.828	1.310	0.053
Gadj 2 ³	0.950	0.805	1.450	0.069
Swedish bulls				
A	1.008	1.000	1.081	0.011
Gorg ²	1.006	0.871	1.184	0.037
Gorg 2 ²	1.006	0.871	1.184	0.037
Gadj ³	0.957	0.775	1.233	0.044
Gadj 2 ³	0.977	0.773	1.364	0.051
Finnish bulls				
A	1.016	1.000	1.135	0.015
Gorg ²	0.979	0.877	1.157	0.028
Gorg 2 ²	0.979	0.877	1.157	0.027
Gadj ³	0.939	0.784	1.283	0.036
Gadj 2 ³	0.961	0.776	1.401	0.043

¹Summaries are across populations and within bulls born in Denmark, Sweden, and Finland.

²Original method (original **G**) 1 and 2 of VanRaden (2008) calculated using allele frequencies observed across breeds.

³Adjusted method (adjusted **G**) 1 and 2 of VanRaden (2008) computed using breedwise observed allele frequencies.

Table 4 shows summaries of diagonal elements from the **G** matrices calculated using estimated base population AF. Variability in the diagonals of **G** matrices calculated using base population AF was lower than when using the currently genotyped population AF. The variability in terms of standard deviation of diagonal elements was 22 and 35% less in the adjusted matrices **Gadj** and **Gadj**2, respectively, for across breeds, and in the range between 17 and 37% for within breeds. The averages of diagonal elements were less than 1 in all populations for the proposed adjusted matrices, irrespective of AF used to calculate **G**. For **Gorg**, a larger mean may have resulted from having, in general, higher diagonal elements. In all cases, the tendencies observed for diagonal elements were also clear for pairwise relationships (results not shown).

Figure 1 shows the distributions of **G**_{ii} from different **G** matrices calculated using observed AF in the combined population. The distributions were examined to ensure consistency with the statistics presented above, as measures such as minimum and maximum tend to be sensitive to extreme values. The shape of the density plot generally followed a normal distribution as suggested with a theoretical example by Simeone et al. (2011). However, tails for all distributions were slightly longer to the right of the distributions, which could

be partly associated with the distribution of **A**_{ii}. The distributions from adjusted matrices (i.e., **Gadj** and **Gadj**2) had only 1 peak, whereas that of **Gorg** appeared to be bimodal. This agrees with the suggestion that multiple peaks would not occur in the distribution of **G**_{ii} within a population, but may be expected in multiple populations if **G** is scaled with AF across breeds (Simeone et al., 2011). The density plots for diagonal elements from **G** were also examined within populations (plots not shown). The distributions for all **G**_{ii} were generally distributed normally in SWE and FIN bulls; however, **Gorg** had a bimodal distribution in the DNK bulls.

Figure 2 shows the distributions of **G**_{ii} from different **G** matrices calculated using estimated base population AF in the combined population. Similarly here as for observed AF, all plots were approximately normally distributed, with slight tails to the right. The bimodal density plot of **Gorg** diagonals was slightly smoothed when AF were estimated from the base population. Plots within populations were also normally distributed in the SWE and FIN bulls whereas **Gorg** appeared bimodal for the DNK animals. The variability of **G**_{ii}, especially for adjusted matrices, was much less when AF were estimated from the base population (Figure 2) than from the current population (Figure 1). In this ad-

Table 4. Descriptive statistics of diagonal elements of the original and adjusted genomic relationship matrices (**G**) estimated using allele frequencies in the base population, across populations, and within bulls born in Denmark, Sweden, and Finland

Item	Mean	Minimum	Maximum	SD
Across populations				
Gorg ¹	1.023	0.906	1.329	0.049
Gadj ²	0.981	0.819	1.224	0.035
Gadj2 ²	0.986	0.824	1.262	0.036
Danish bulls				
Gorg ¹	1.088	0.951	1.275	0.057
Gadj ²	0.994	0.868	1.204	0.037
Gadj2 ²	0.976	0.833	1.226	0.044
Swedish bulls				
Gorg ¹	1.008	0.906	1.152	0.029
Gadj ²	0.980	0.869	1.137	0.032
Gadj2 ²	0.986	0.887	1.166	0.032
Finnish bulls				
Gorg ¹	1.006	0.935	1.147	0.026
Gadj ²	0.975	0.871	1.184	0.030
Gadj2 ²	0.989	0.868	1.199	0.032

¹Original method (original **G**) 1 of VanRaden (2008) calculated using base population allele frequencies across breeds.

²Adjusted method (adjusted **G**) 1 and 2 of VanRaden (2008) computed using breedwise allele frequencies from the base population.

mixed population, the correlations between **A**_{ii} and **G**_{ii} by any of the methods were always close to zero when using population-level AF but increased to 0.16, 0.28, and 0.38, respectively, for **Gorg**, **Gadj**, and **Gadj2**

when AF were estimated from the base population. The correlations within populations were also higher and ranged from 0.26 for **Gadj** to 0.53 for **Gorg** when AF were estimated from the base populations.

Figure 1. The distributions of diagonal elements from genomic matrices (**G**) calculated using estimated allele frequencies (AF) in the currently genotyped population. **Gorg** = original method 1 (original **G**) of VanRaden (2008), calculated using AF across breeds. **Gadj** and **Gadj2** = new adjusted methods (adjusted **G**) 1 and 2 of VanRaden (2008), respectively, computed using breedwise observed AF.

Validation Reliabilities of DGV

Table 5 shows the regression coefficients and validation reliabilities for milk, protein, and fat obtained using alternative **G** matrices. The DGV were from **G** matrices calculated using AF in the current and the base population only because AF estimated from genotypes of old individuals appeared to be less usable. The regression coefficients and validation reliabilities were similar for all matrices irrespective of whether breedwise or across-breed AF were used and whether AF were estimated in the current or base population. Thus, predictions of genomic values converged to similar solutions regardless of AF used to compute **G**. For all matrices, regression coefficients were less than the expected value of 1 for milk (0.71), protein (0.75), and fat (0.81). The validation reliabilities were low for milk (0.33) and protein (0.33) and slightly higher for fat (0.43).

DISCUSSION

An important step when defining the model relates to the genetic covariance between relatives, which reflects shared genes that arise through common ancestry. High-density panels of SNP markers have recently been used to estimate genomic relationships in addition

Figure 2. The distributions of diagonal elements from genomic matrices (**G**) calculated using estimated allele frequencies (AF) in the base population. **Gorg** = original method 1 (original **G**) of VanRaden (2008), calculated using AF across breeds. **Gadj** and **Gadj2** = new adjusted methods (adjusted **G**) 1 and 2 of VanRaden (2008), respectively, computed using breedwise base population AF.

to the traditional pedigree-based relationships. Several methods for the estimation of **G** within a breed have been proposed in the literature (VanRaden, 2008; Misztal et al., 2009; Christensen and Lund, 2010; Yang et al., 2010). A general agreement across studies about genomic selection is that there is a gain in prediction accuracies for young unproven bulls when genomic information is incorporated compared with traditional evaluations with pedigree information only, due to improved prediction of Mendelian sampling deviations between close relatives in **G**. Methods for calculating **G** are straightforward in a single-breed population;

however, similar approaches tend to result in distorted coefficients in multibreed populations. Thus, relationships should account for the different expectations for mean and variance, depending on breed composition of the individuals in multibreeds (Harris and Johnson, 2010). Only a few studies have evaluated the prospect of accounting for multibreeds in **G** with real data (Harris and Johnson, 2010; VanRaden et al., 2011). The resulting **G** has been tested in genomic predictions for crossbred animals in New Zealand (Harris and Johnson, 2010). The objectives of this study were to examine the prospects of accounting for breed composition in the calculation of **G** and assess coefficients of different **G** matrices compared with the pedigree-based relationship matrix within and across breeds in a multibreed population.

Breedwise AF

Allele frequencies play a crucial role in the calculation of **G** and, hence, erroneous estimation of AF may result in biased **G** coefficients. Our rationale behind the estimation of breedwise AF was that individuals from breeds that developed independently would likely have different AF. We found that the approach proposed by Gengler et al. (2007) for estimating gene content of ungenotyped individuals given pedigree useful in the estimation of breedwise AF in the base population. The fewer number of markers with AF outside the expected range of 0 and 1 found when AF were estimated in the base population indicates that the pedigree was better able to differentiate base breeds compared with AF estimated in the currently genotyped animals. Estimates of AF outside the expected range result from using a simplified model without restrictions on the parameter space. Restricting the parameter space to fall between 0 and 1 using a binomial model resulted in coefficients close to these values. The AF out of bounds did not create a great problem in the current population, as no pure base breed animals were included and, therefore,

Table 5. Regression coefficients (b_1) and validation reliabilities of direct estimated genomic values (R^2_{DGV}) from genomic relationship matrices (**G**) calculated using currently genotyped and base population allele frequencies (AF)¹

Trait	Observed AF						Base population AF					
	Gorg		Gadj		Gadj2		Gorg		Gadj		Gadj2	
	b_1	R^2_{DGV}	b_1	R^2_{DGV}	b_1	R^2_{DGV}	b_1	R^2_{DGV}	b_1	R^2_{DGV}	b_1	R^2_{DGV}
Milk	0.71	0.32	0.71	0.32	0.72	0.33	0.71	0.32	0.71	0.32	0.72	0.33
Protein	0.75	0.33	0.75	0.33	0.76	0.33	0.75	0.33	0.75	0.33	0.76	0.33
Fat	0.81	0.43	0.80	0.42	0.82	0.43	0.81	0.43	0.80	0.42	0.82	0.43

¹**Gorg** = original method (original **G**) 1 of VanRaden (2008) calculated using AF across breeds; **Gadj** and **Gadj2** = new adjusted methods (adjusted **G**) 1 and 2 of VanRaden (2008), respectively, computed using breedwise AF.

an individuals' expectation of AF was generally correct. However, if purebreds are included in the data, their expected AF may be imprecise; it may be useful in this case to use a binomial model. An alternative approach to Gengler's fixed breed effects would be to include unknown parent groups for each base breed in the matrix $\mathbf{A}^{-1}$ (Quaas, 1988). This model would yield genetic group effects equivalent to our AF within breed in the base population.

The correlations between estimated breedwise AF were also higher using the gene content approach. When considering only the minor AF, correlations dropped but were still higher in the base population. The observed high correlations between breedwise AF estimated from the base population compared with low correlations between these frequencies in the currently genotyped population have been reported in VanRaden et al. (2011). The authors pointed out that this indicates that recent drift within populations was removed during the estimation process and also revealed that breeds had been more similar over 10 generations in the past than they are at present, as expected from genetic drift of frequencies across pedigree generations. This progressive differentiation in AF between breeds, however, may not have been the case for the current population, as base breeds were combined and subjected to the same breeding goal for over 2 decades, which is expected to make them similar genetically. The higher correlations found between SRB, FAY, and the breed "other" were unusual. However, many animals with SRB and FAY fractions also have a breed proportion for Canadian Ayrshires that is now in the breed "other." Thus, the estimation of AF might have detected relations to such animals. The expected pedigree-based breed proportions were used in this study to define the 4 breeds. Alternatively, accurate prediction of breed composition based on SNP genotype data has been demonstrated for multibreed populations (Kuehn et al., 2011; Frkonja et al., 2012). However, these algorithms initially estimate breed-specific frequencies using purebred individuals, which was a limitation in this population due to unavailability of such individuals.

Properties of $\mathbf{A}$ and $\mathbf{G}$ Within and Across Populations

The diagonal elements of all $\mathbf{G}$ matrices that used observed AF were incomparable to the diagonal elements of $\mathbf{A}$. Moreover, the variability in $\mathbf{A}$ was much smaller than observed in $\mathbf{G}$ matrices. The comparison of $\mathbf{A}$ and $\mathbf{G}$ computed using observed AF is generally vague, as the coefficients in $\mathbf{A}$ are expressed relative to the base population in the distant past and the additive genetic variance is defined for that generation. In contrast, when the base population for $\mathbf{G}$ is achieved

by scaling IBD coefficients with observed AF, the additive genetic variance among animals considers average variation in the current genotyped animals (Powell et al., 2010; Yang et al., 2010). The estimate of additive genetic variance may be smaller in the current population than it was in the distant past because we expect the current population to be more inbred (Powell et al., 2010). The highest variability in diagonal elements from **Gorg** and **Gorg2** indicates that the use of across-breed AF increased coefficients in $\mathbf{G}$ for this admixed population. Variability was reduced by using breedwise AF in **Gadj**. However, in **Gadj** and **Gorg**, the overall scaling was based on the same marker variance across breeds, which was larger than the expected variance within breeds. Consequently, the variance of the diagonal elements appeared much smaller in **Gadj** than in **Gorg**. The simplified scaling factor was corrected in **Gadj2**, where elements were scaled by the mean marker variance of an individual's base breeds. The variance from the resulting **Gadj2** was generally still smaller than observed from the original approaches across breeds in **Gorg** and **Gorg2**.

Diagonal elements of $\mathbf{G}$ built using any of the approaches were more accurately estimated when the $\mathbf{G}$ used AF estimated from the base population. The variability of diagonal elements, particularly for **Gadj2**, was reduced to a greater extent, suggesting that breedwise AF in this case were less biased and highlights a great need to account for the pedigree structure in $\mathbf{G}$ for multibreed populations. Thus, with pedigree information, the base (founder) population is generally consistent. The calculated diagonals of $\mathbf{G}$ in this case were moderately correlated with $\mathbf{A}$ across breeds and in agreement with that reported by Aguilar et al. (2010) in the US Holstein population. However, the moderate correlation of 0.38 between diagonals of $\mathbf{A}$ and **Gadj2** calculated with base population AF in our study was much lower than correlation of 0.68 reported with simulated data (VanRaden, 2008) and correlations ranged from 0.50 to 0.56 for the US Holstein, Jersey, and Brown Swiss populations (VanRaden et al., 2011). The highest correlations were generally found with true AF in simulated data (VanRaden, 2008) and when all AF were assumed to be 0.5 for all markers (Hayes and Goddard, 2008; VanRaden et al., 2011). Assuming AF equal to 0.5 for all markers equalizes the relative contribution of markers to $\mathbf{G}$ instead of having rare alleles contributing more than common alleles (Forni et al., 2011). In addition to AF, one reason why these correlations were different across studies might be due to different population structures. The advantage of base breedwise AF in **Gadj2** could mean that the distant past founder population corresponds well with the current expected homozygosity. Moreover, our results suggest that calcu-

lation of $\mathbf{G}$ with respect to an individual's base breed corrected better the heterogeneity than simple animal deviations from across-population mean AF.

When $\mathbf{G}$ coefficients were assessed within populations, we found that using across-breed AF tended to increase $\mathbf{G}_{ii}$ for animals from populations that had fewer animals or that were distantly related to dominating breeds in the combined population. Simeone et al. (2011) indicated that with an equal number of animals contributing to the AF across populations, there would be fewer differences in scaling $\mathbf{G}$ between populations. The mean AF across breeds was strongly influenced by the Swedish and Finnish populations, as these breeds are more related genetically (Brøndum et al., 2011; Makgahlela et al., 2013) and had more animals in the combined data. As a result, and in contrast to $\mathbf{A}_{ii}$, the general level of homozygosity in the Danish population appeared to be higher than in other populations. This is unexpected because the Danish population has been found to be more admixed than the other 2, due to years of crossbreeding (Brøndum et al., 2011; Makgahlela et al., 2013). This means that as diagonal elements from $\mathbf{G}_{org}$ have been increased in the Danish bulls, they were decreased for animals in the other populations. In addition, individuals with the highest diagonal elements in $\mathbf{G}_{org}$ were found to be registered elsewhere but not in the Nordic RDC. Because these animals come from populations with AF deviating even further from the population mean AF, their genotypes make them appear more homozygous than the average homozygosity in this population. Apart from a great reduction in the variability of diagonal elements between AF estimated from the current and base population, the behaviors of different estimators of $\mathbf{G}$ were similar within and across populations.

Harris and Johnson (2010) indicated that multibreed reference populations will lead to biased coefficients in $\mathbf{G}$ if breed is not taken into account. In this study, the use of breedwise AF to calculate $\mathbf{G}_{adj}$ and $\mathbf{G}_{adj2}$ reduced country differences in coefficients similarly within and across populations. Regarding distributions, the observation that diagonal elements of $\mathbf{G}_{adj}$ and $\mathbf{G}_{adj2}$ generally followed a normal distribution, but diagonal elements of $\mathbf{G}_{org}$ appeared bimodal, indicates a distortion in the elements of $\mathbf{G}_{org}$ and suggests clusters that may be due to the population structure. This observation was in agreement with previous findings (Simeone et al., 2011). In their study, the authors used simulated data on multiple populations and 60,000 SNP markers with varying AF at each locus to compute $\mathbf{G}$ using observed AF across populations. They observed a bimodal distribution of the diagonal elements of $\mathbf{G}$. Multiple peaks were correctly avoided by using breedwise AF in our study. Simeone et al. (2011) pointed

out a general lack of theoretical knowledge about the distribution of the diagonal elements of $\mathbf{G}$ both within and across breeds.

The estimation of relationships and their use in predictions is widely carried out within a breed or treating multibreed data as a homogeneous breed (Hayes et al., 2009; Pryce et al., 2011; Su et al., 2012), except in New Zealand where predictions account for breed effects for crossbred animals (Harris and Johnson, 2010). The observed correlations between diagonal elements from $\mathbf{A}$ and $\mathbf{G}$ within breed were higher when estimated base population AF were used instead of current population AF (VanRaden 2008; VanRaden et al., 2011) or with AF equal to 0.5 (VanRaden et al., 2011). Our use of estimated breedwise AF also greatly improved diagonal elements, indicating that AF have a large effect on relationship coefficients and would also have an effect on the estimation of population additive genetic variation. Using pig data, Forni et al. (2011) explored different methods of scaling $\mathbf{G}$ with observed AF. They estimated genetic variances using different $\mathbf{G}$ for genotyped animals only, and estimates ranged from 2.25 when $\mathbf{G}_{ii}$ were normalized to average to 1 and were inflated to 4.46 when $\mathbf{G}$ was scaled with the expectations of AF following a β distribution (Gianola et al., 2009). The estimated additive genetic variances were more sensitive when a selected subset of genotyped animals was used for variance components estimation. Similar additive genetic variances were found when complete data of genotyped and ungenotyped animals were used (Forni et al., 2011).

In the absence of selective genotyping, the expected regression coefficient (b_1) from the validation model is close to 1 (Mäntysaari et al., 2010). The observed b_1 values (range = 0.71–0.82) were less than the expected value of 1, which indicates bias in the estimated genomic values. The validation reliabilities of DGV from all $\mathbf{G}$ matrices were similar. In all cases, solutions indicated that DGV were unaffected by the AF used to calculate $\mathbf{G}$. This observation agrees with previous reports where the gain in accuracy of DGV was small (0.01) when base population AF were used to compute $\mathbf{G}$ instead of observed AF in simulated data (VanRaden, 2008), and were indifferent to AF with real data (Forni et al., 2011). According to Strandén and Christensen (2011), DGV are neither sensitive to marker allele coding nor AF and the same DGV solutions would be calculated in GBLUP, provided that the model has a common fixed general mean effect. Thus, the absolute levels of values (i.e., animal effects) are only affected when the mean is uncounted for in the model. Similarly here, inclusion of fixed-breed regressions for $\mathbf{G}_{adj2}$ brought breed means back into the DGV. It was clearly shown that different AF affected the calculation of $\mathbf{G}$. Although

G was sensitive to AF and was accurately computed using breedwise AF in the base population, the DGV validation reliabilities were indifferent. In multibreed populations, the use of **Gadj2** may be more beneficial in single-step evaluations where most animals are evaluated by matrix **A** and through their relationships to genotyped animals. However, it should be emphasized that the use of across-breed MAF at least 5% to select SNP tend to remove markers within breed that may be informative for improved prediction accuracy. In the presence of purebred animals, it may be useful to select SNP based on MAF of 1 breed, even if monomorphic in all other breeds (Olson et al., 2012). An alternative to adjusting breeds in **G** would be to estimate SNP effects in different breeds simultaneously in a multi-trait regression model (Olson et al., 2012; Makgahlela et al., 2013). The prediction accuracies were found to be even higher when fitting correlated SNP effects between breeds (Olson et al., 2012). The unavailability of purebred animals in the data also limited comparisons of our expected breedwise AF from the actual estimates within breed.

CONCLUSIONS

Current methods used for computing genomic relationships in multibreed populations need to be extended to allow for differential AF between breeds. This study showed that errors in the estimation of AF may have great consequences in the calculation of relationships. Across-breed observed AF increased diagonal elements of **G** for animals from breeds that are distantly related to the combined population and have fewer animals in the combined population. Breedwise AF reduced country differences in **G** similarly within and across populations, resulting in a normal distribution of diagonal elements. Breedwise AF were more accurately estimated when accounting for the pedigree structure or estimated from the base population, thereby reducing the variability of diagonal elements of **Gadj2**. The DGV and their validation reliabilities were unaffected by AF used to compute **G** when estimated using a GBLUP model. The method for **Gadj2** may provide more consistency among relationship coefficients between genotyped and ungenotyped individuals in an across-breed single-step evaluation.

ACKNOWLEDGMENTS

The authors acknowledges the Nordic Cattle Genetic Evaluation Ltd. (Aarhus, Denmark) and Nordic Genomic Selection project for providing the genotype and phenotype data. M. L. Makgahlela acknowledges financial support from the Finnish Ministry of Agriculture

and Forestry (Helsinki, Finland) and the University of Helsinki in Finland.

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.
- Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtssen, W. F. Fikse, and M. S. Lund. 2011. Reliabilities of genomic prediction using combined reference data of the Nordic red dairy cattle populations. *J. Dairy Sci.* 94:4700–4707.
- Christensen, O. F., and M. S. Lund. 2010. Genomic prediction when some animals are not genotyped. *Genet. Sel. Evol.* 42:2.
- Eding, H., and T. H. E. Meuwissen. 2001. Marker-based estimates of between and within population kinships for the conservation of genetic diversity. *J. Anim. Breed. Genet.* 118:141–159.
- Fernández, J., T. H. E. Meuwissen, M. A. Toro, and A. Mäki-Tanila. 2011. Management of genetic diversity in small farm animal populations. *Animal* 5:1684–1698.
- 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.
- Frkónja, A., B. Gredler, U. Schnyder, I. Curik, and J. Sölkner. 2012. Prediction of breed composition in an admixed cattle population. *Anim. Genet.* 43:696–703.
- Gengler, N., P. Mayeres, and M. Szydlowski. 2007. A simple method to approximate gene content in large pedigree populations: Application to the myostatin gene in dual-purpose Belgian blue cattle. *Animal* 1:21–28.
- Gianola, D., G. de los Campos, W. G. Hill, E. Manfredi, and R. Fernando. 2009. Additive genetic variability and the Bayesian alphabet. *Genetics* 183:347–363.
- Gilmour, A. R., B. J. Gogel, B. R. Cullis, and R. Thompson. 2009. ASREML User Guide Release 3.0. VSN International Ltd., Hemel Hempstead, UK.
- Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* 177:2389–2397.
- Harris, B. L., and D. L. Johnson. 2010. Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *J. Dairy Sci.* 93:1243–1252.
- Hayes, B. J., and M. E. Goddard. 2008. Technical note: Prediction of breeding values using marker-derived relationship matrices. *J. Anim. Sci.* 86:2089–2092.
- Hayes, B. J., P. M. Visscher, and M. E. Goddard. 2009. Increased accuracy of artificial selection by using the realized relationship matrix. *Genet. Res. (Camb.)* 91:47–60.
- Illumina Inc. 2005. Illumina GenCall Data Analysis Software-Gen-Call software algorithms for clustering, calling, and scoring genotypes. Illumina. Pub. No. 370-2004-009. Illumina Inc., San Diego, CA.
- Kuehn, L. A., J. W. Keele, G. L. Bennett, T. G. McDanel, T. P. L. Smith, W. M. Snelling, T. S. Sonstegard, and R. M. Thallman. 2011. Predicting breed composition using breed frequencies of 50,000 markers from the US meat animal research center 2,000 bull project. *J. Anim. Sci.* 89:1742–1750.
- Legarra, A., I. Aguilar, and I. Misztal. 2009. A relationship matrix including full pedigree and genomic information. *J. Dairy Sci.* 92:4656–4663.
- Lidauer, M., E. A. Mäntysaari, I. Strandén, J. Pösö, J. Pedersen, U. S. Nielsen, K. Johansson, J.-Å. Eriksson, P. Madsen, and G. P. Aamand. 2006. Random heterosis and recombination loss effects in a multibreed evaluation for Nordic Red dairy cattle. *Proc. 8th World Congr. Genet. Appl. Livest. Prod., Belo Horizonte, Brazil.*
- Lidauer, M., and I. Strandén. 1999. Fast and flexible program for genetic evaluation in dairy cattle. *Interbull Bull.* 20:20.
- Makgahlela, M. L., E. A. Mäntysaari, I. Strandén, M. Koivula, U. S. Nielsen, M. J. Sillanpää, and J. Juga. 2013. Across breed multi-

- trait random regression genomic predictions in the Nordic Red dairy cattle. *J. Anim. Breed. Genet.* 130:10–19.
- Malécot, G. 1948. *Les Mathématiques de L'hérédité*. Masson et Cie, Paris, France.
- Mäntysaari, E. A., M. Koivula, I. Strandén, J. Pösö, and G. P. Aamand. 2011. Estimation of GEBVs using deregressed individual cow breeding values. *Interbull Bull.* 44:26–29.
- Mäntysaari, E. A., Z. Liu, and P. VanRaden. 2010. Interbull validation test for genomic evaluations. *Interbull Bull.* 41:17–22.
- Misztal, I., A. Legarra, and I. Aguilar. 2009. Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *J. Dairy Sci.* 92:4648–4655.
- Mrode, R. A., and G. J. T. Swanson. 2004. Calculating cow and daughter yield deviations and partitioning of genetic evaluations under a random regression model. *Livest. Prod. Sci.* 86:253–260.
- Olson, K. M., P. M. VanRaden, and M. E. Tooker. 2012. Multibreed genomic evaluations using purebred Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 95:5378–5383.
- 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.
- Pryce, J. E., B. Gredler, S. Bolormaa, P. J. Bowman, C. Egger-Danner, C. Fuerst, R. Emmerling, J. Sölkner, M. E. Goddard, and B. J. Hayes. 2011. Short communication: Genomic selection using a multi-breed, across-country reference population. *J. Dairy Sci.* 94:2625–2630.
- Quaas, R. L. 1988. Additive genetic model with groups and relationships. *J. Dairy Sci.* 71:1338–1345.
- Scheet, P., and M. Stephens. 2006. A fast and flexible statistical model for large-scale population genotype data: Applications to inferring missing genotypes and haplotypic phase. *Am. J. Hum. Genet.* 78:629–644.
- Simeone, R., I. Misztal, I. Aguilar, and A. Legarra. 2011. Evaluation of the utility of diagonal elements of the genomic relationship matrix as a diagnostic tool to detect mislabelled genotyped animals in a broiler chicken population. *J. Anim. Breed. Genet.* 128:386–393.
- Strandén, I., and O. F. Christensen. 2011. Allele coding in genomic evaluation. *Genet. Sel. Evol.* 43:25.
- Strandén, I., and K. Vuori. 2006. RelaX2: Pedigree analysis programme. *Proc. 8th World Congress Genetics Applied Livest. Prod.*, Belo Horizonte, Brazil.
- Su, G., P. Madsen, U. S. Nielsen, E. A. Mäntysaari, G. P. Aamand, O. F. Christensen, and M. S. Lund. 2012. Genomic prediction for Nordic Red cattle using one-step and selection index blending. *J. Dairy Sci.* 95:909–917.
- Toro, M. A., T. H. E. Meuwissen, J. Fernández, I. Shaat, and A. Mäkitanila. 2011. Assessing the genetic diversity in small farm animal populations. *Animal* 5:1669–1683.
- VanRaden, P. M. 2008. Efficient methods to compute genomic predictions. *J. Dairy Sci.* 91:4414–4423.
- VanRaden, P. M., K. M. Olson, G. R. Wiggans, J. B. Cole, and M. E. Tooker. 2011. Genomic inbreeding and relationships among Holsteins, Jerseys, and Brown Swiss. *J. Dairy Sci.* 94:5673–5682.
- 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.
- Zimin, A. V., A. L. Delcher, L. Florea, D. R. Kelley, M. C. Schatz, D. Puiu, F. Hanrahan, G. Pertea, C. P. van Tassell, T. S. Sonstegard, G. Marçais, M. Roberts, P. Subramanian, J. A. Yorke, and S. L. Salzberg. 2009. A whole-genome assembly of the domestic cow, *Bos taurus*. *Genome Biol.* 10:R42.