J. Dairy Sci. 94:4700–4707
doi:10.3168/jds.2010-3765

Reliabilities of genomic prediction using combined reference data of the Nordic Red dairy cattle populations

R. F. Brøndum,*[1] E. Rius-Vilarrasa,† I. Strandén,‡ G. Su,* B. Guldbrandtsen,* W. F. Fikse,† and M. S. Lund*
*Aarhus University, Faculty of Science and Technology, Department of Genetics and Biotechnology, DK-8830 Tjele, Denmark
†Swedish University of Agricultural Sciences, Department of Animal Breeding and Genetics, PO Box 7023, 750 07 Uppsala, Sweden
‡MTT, Biometrical Genetics, 31600 Jokioinen, Finland

ABSTRACT

This study investigated the possibility of increasing the reliability of direct genomic values (DGV) by combining reference populations. The data were from 3,735 bulls from Danish, Swedish, and Finnish Red dairy cattle populations. Single nucleotide polymorphism markers were fitted as random variables in a Bayesian model, using published estimated breeding values as response variables. In total, 17 index traits were analyzed. Reliabilities were estimated using a 5-fold cross validation, and calculated as the within-year squared correlation between estimated breeding values and DGV. Marker effects were estimated using reference populations from individual countries, as well as using a combined reference population from all 3 countries. Single-country reference populations gave mean reliabilities across 17 traits of 0.19 to 0.23, whereas the combined reference gave mean reliabilities of 0.26 for all populations. Using marker effects from 1 population to predict the other 2 gave a loss in mean reliability of 0.14 to 0.21 when predicting Swedish or Finnish animals with Danish marker effects, or vice versa. Using Swedish or Finnish marker effects to predict each other only showed a loss in mean reliability of 0.03 to 0.05. A combined Swedish-Finnish reference population led to an average reliability as high as that from the 3-country reference population, but somewhat different for individual traits. The results from this study show that it is possible to increase the reliability of DGV by combining reference populations from related populations.
Key words: combined reference population, genomic selection, reliability

INTRODUCTION

Genomic selection is becoming a popular tool in cattle breeding. It is based on breeding values predicted directly from dense sets of genetic markers, possibly combined with pedigree information. Each genetic marker from an SNP panel is potentially in linkage disequilibrium (**LD**) with QTL. The effect of each allele at the individual marker loci is estimated by fitting a model to phenotypic data from a group of reference animals that have both genomic and phenotypic records (or pseudo-observations). The genomic breeding value of a candidate with no phenotypic record is then predicted as the sum of all marker effects (Meuwissen et al., 2001).

Previous studies have shown that the reliability of direct genomic values (**DGV**) largely depends on the number of animals used to determine the marker effects as well as the heritability of the trait (Goddard and Hayes, 2009). VanRaden et al. (2009) reported an increase in the reliability of DGV when increasing the size of the reference data in a Holstein population. These findings suggest that for a population having a small number of reference animals, the reliabilities of DGV might be lower than the reliabilities of the parent averages. However, Hayes et al. (2009b) showed that it is possible to increase the reliability of DGV in Australian Jersey cattle by including Australian Holstein cattle in the reference data. The increase was, however, small and not present for all traits. Su et al. (2009) reported a considerable improvement in genomic prediction in a simulation study combining reference data from 2 populations having a common origin. Results from the latter study also show that the weaker the genetic ties between the populations, the smaller the gain in reliability of DGV when combining them.

To date, reliabilities of DGV have been reported in Holstein, Jersey, and Norwegian Red cattle

(Harris et al., 2008; Hayes et al., 2009a; Luan et al., 2009; VanRaden et al., 2009). However, the reliability of genomic prediction in the Danish, Swedish, and Finnish Red dairy cattle populations has not yet been investigated. These 3 populations have strong genetic ties due to some bulls in common use. Especially the Swedish and Finnish populations share much of their genetic origin, whereas the Danish population also has influences from Danish Red Holstein and American Brown-Swiss cattle (Team Avlsværdivurdering, 2009).

Received August 27, 2010.
Accepted May 5, 2011.
[1]Corresponding author: rasmusf.brondum@agrsci.dk

Table 1. Traits selected for the study: size of reference population (ref), mean and standard deviation of EBV, mean reliability of the EBV, and reliability of parent average (PA) at the time of birth

Trait	Size of ref	Mean EBV	SD of EBV	Mean reliability	Reliability of PA
Milk yield[1]	3,550	93.68	10.77	94.12	0.38
Fat yield[1]	3,550	93.90	10.85	94.12	0.38
Protein yield[1]	3,550	91.71	12.22	94.12	0.38
Yield index[1]	3,550	91.76	12.25	94.12	0.38
Udder health[2]	3,785	98.20	10.20	80.72	0.25
Fertility[2]	3,755	99.77	10.10	78.39	0.22
Other diseases[2]	3,705	104.59	10.93	64.11	0.18
Longevity[2]	3,556	97.90	8.14	59.28	0.21
Direct calving[2]	3,776	99.81	10.45	77.93	0.26
Maternal calving[2]	3,778	98.37	9.47	68.52	0.22
Body[1]	2,743	98.99	10.90	81.71	0.30
Feet and legs[1]	3,095	97.45	8.53	62.20	0.26
Udder conformation[1]	3,095	96.12	9.93	77.74	0.33
Milking ability[1]	3,095	97.44	11.59	81.53	0.25
Temperament[1]	3,095	96.56	10.42	74.06	0.25
Nordic total merit index[1]	3,785	−10.39	12.84	—	—
Growth[1]	1,901	95.68	11.26	90.36	0.28
Mean	—	—	—	79.56	0.32[1] 0.22[2]

[1]Predicted using animal model.

[2]Predicted using sire model.

Separately, each of the 3 populations have small reference populations, which might result in low reliabilities of predicted DGV. Because of the strong genetic ties among the 3 populations, an increase in the reliability of DGV can be expected when combining the reference populations.

In the present study, we investigated the efficiency of genomic prediction using 3 approaches. First, DGV were obtained separately for each of the 3 national populations. Second, the possibility of increasing the reliability of genomic prediction by pooling reference data from the Danish, Swedish, and Finnish Red dairy cattle populations was investigated. In addition, a combined Swedish-Finnish reference population was also investigated. Finally, we investigated the efficiency of using a reference population from one country to predict breeding values of animals from another national population. This will be referred to as the cross-prediction study.

MATERIALS AND METHODS

Data

The data set included 778 Danish, 1,395 Swedish, and 1,562 Finnish Red dairy bulls from 306 half-sib families, born between 1986 and 2005. The bulls were genotyped using Illumina Bovine SNP50 BeadChip (Matukumalli et al., 2009). Phenotypic data were conventional EBV evaluated in 2009 by the Nordic Cattle Genetic Evaluation, where a joint Nordic model is routinely used to predict EBV for the 3 national populations. The Nordic total merit index and its 16 sub-index traits (see Table 1) were selected for this study. A detailed description of the traits can be found by Danish Cattle Federation (2006).

Three separate national studies were conducted for the Danish, Swedish, and Finnish bulls. In the Danish study, Swedish animals from Danish/Swedish half-sibling families were included in the data set if the number of Swedish bulls in the family was less than 5 times the number of Danish bulls. A similar approach was applied in the Swedish study for including Danish animals. In the Finnish study and in the cross predictions, only animals of the same nationality were used as a reference.

Editing of Genotypic Data

The genotypic data was edited both by animal and by loci. The number of markers used in the different reference populations is shown in Table 2. For animals, the requirements were a call rate above 95% except for some old animals, which were accepted with call rates of at least 85%. These older animals were accepted with lower call rates because they were important bulls, where the quality of the DNA was poor due to extended storage. Marker loci were accepted if they had a call rate of at least 95% in a large reference sample of Danish Holstein bulls. Loci with a minor allele frequency less than 5% were excluded. Loci without a map position in the Btau 4.0 assembly were discarded. Animals with an average GenCall score (Illumina Inc., 2005) of less than

Table 2. Number of markers used in the predictions for Danish (DK), Swedish (SWE), Finnish (FIN), Swedish-Finnish (SWE-FIN), combined (COMB), and cross-prediction (Cross) reference populations

Item	DK	SWE	FIN	SWE-FIN	COMB and Cross
Number of markers	38,764	38,293	31,890	38,315	38,755

0.65 were discarded. Individual marker typings with a GenCall score of less than 0.6 were discarded. In the Swedish, Finnish, and combined Swedish-Finnish studies, the X chromosome was omitted from the data set. Comparison of reliability for a few traits done with and without the X chromosome, however, shows no significant changes (results not shown). For the national studies, selection of marker based on minor allele frequency was done based on calculations from a reference sample of Danish Holstein bulls. In the Finnish study, marker frequencies were, however, recalculated, and further markers removed, which explains the lower number of markers in this population.

Statistical Model

In this study, SNP markers were used as predictors and EBV as response variables. The EBV were weighted using a function of the reliability of the EBV given by 1/(1 – reliability of EBV) and scaled to a mean weight of 1. Bayesian inference was used to estimate the SNP effects, using the model

$$\mathbf{y} = \mathbf{1}\mu + \sum_{i=1}^{m} \mathbf{X}_i \mathbf{q}_i v_i + \mathbf{e},$$

where $\mathbf{y}$ is the vector of conventional EBV, μ is the intercept, m is the number of SNP markers, $\mathbf{X}_i$ is the design matrix allocating alleles at locus i to the animals, $\mathbf{q}_i$ is the vector of scaled marker effects at locus i, v_i is a scaling factor at locus i, and $\mathbf{e}$ is the vector of residuals. The prior distributions were

$$\mathbf{q}_i \sim N(\mathbf{0}, \mathbf{I}) \qquad v_i \sim TN(0, \sigma_v^2) \qquad \mathbf{e} \sim N(\mathbf{0}, \mathbf{W}\sigma_e^2),$$

where $\mathbf{I}$ is a 2 × 2 identity matrix, $\mathbf{W}$ is a diagonal matrix containing the inverse of the weights, and TN is a truncated normal distribution. A complete description of the model can be found in Su et al. (2010) and Villumsen et al. (2009).

The DGV for individual k was calculated as

$$DGV_k = \hat{\mu} + \sum_{i=1}^{m} \mathbf{X}_{i,k} \hat{\mathbf{q}}_i \hat{v}_i.$$

Marker effects were estimated using the iBay v 1.46 software package (Luc Janss, Faculty of Science and Technology, Aarhus University, Tjele, Denmark). The Gibbs sampler was run as a single chain with 50,000 iterations. Samples from the first 10,000 iterations were discarded as burn-in, and every fifth sample of the remaining 40,000 iterations was saved to estimate parameters based on their posterior distribution. The DGV were calculated as the posterior means. To validate the sufficiency of convergence and the chain length, an additional analysis for 3 traits (protein yield, fertility, and udder health) was carried out with 100,000 iterations and the first 40,000 as burn-in. Comparison of DGV from different chain lengths (50,000 vs. 100,000) showed a correlation above 0.99. The runtime for each trait was approximately 48 h when using the largest reference population, thus making it feasible for routine evaluation if the traits are analyzed in parallel.

Evaluation of Reliability

Reliabilities in the 3 national studies and in the combined reference sets were assessed using a 5-fold cross validation, where animals were divided into 5 approximately equal-sized subsets according to year of birth. Half-sibling families having animals in more than 1 subset were moved to the subset containing the largest part of the family. Cross validation was done by successively removing 1 subset at a time from the whole data set, and using the left-out subset as a test data set. In the cross-prediction study, however, 1 national population was used as a reference, and the other 2 as test data sets. To diminish dependency between reference and test data, sires that had sons in the reference data were excluded from the test data. Table 3 shows how the cross validation was set up for the combined reference set. The other cross validations were performed in a similar manner. When evaluating reliability for the single nationalities based on the combined references, the test sets were obtained by selecting the subset of either Danish, Swedish, or Finnish animals from all of the 5 validation groups.

The reliability of DGV was estimated as the within-year squared correlation between EBV and DGV in the test populations, to remove effects of genetic trend:

$$R^2_{DGV} = \mathrm{Cor}(\mathbf{EBV} - \mathbf{EBV}_{\mathbf{year}}, \mathbf{DGV} - \mathbf{DGV}_{\mathbf{year}}),$$

where $\mathbf{EBV}$ and $\mathbf{DGV}$ are vectors of EBV and DGV, and $\mathbf{EBV}_{\mathbf{year}}$ and $\mathbf{DGV}_{\mathbf{year}}$ are vectors of the annual means of EBV and DGV.

Table 3. Cross-validation groups for all bulls used in the combined reference set[1]

Group	Birth year of family	Families (no.)	Bulls (no.)	Birth year of bulls
1	1986–1990	85	820	1986–1995
2	1991–1994	61	839	1990–1998
3	1995–1997	50	745	1994–2002
4	1998–2000	46	634	1992–2002
5	2001–2005	64	697	1999–2005
Total	—	306	3,735	—

[1]Birth year of family is the criterion that was used for dividing the groups. The birth year of bulls is the actual span of birth years for bulls in the group after animals were moved around to prevent families from overlapping more than 1 group.

RESULTS AND DISCUSSION

Results from the national studies and the cross prediction study are shown in Table 4. The national studies show mean reliabilities between 0.19 and 0.23. The cross predictions show that a Danish reference leads to a loss in mean reliability of 0.14 and 0.17 when predicting the Swedish and Finnish animals, respectively, in comparison to references of the same nationality. Vice versa, a Swedish or Finnish reference causes a loss of 0.18 to 0.21 in mean reliability when predicting the Danish animals. This is in accordance with results from previous studies on using reference data from 1 population to predict DGV for animals from other populations. On the other hand, the Finnish and Swedish populations predict each other with a loss in reliability of only 0.03 to 0.05, compared with results using reference data from the same population. This supports stronger genetic links between the Swedish and Finnish populations than between the Danish and the Swedish or Finnish populations, as suspected from their breeding history. This is confirmed by looking at the genomic relationship matrix (VanRaden, 2008), as shown in Figure 1, and the persistence of LD phase between the 3 populations. Correlation of LD phase in the interval 1 to 700kb was found to be 0.55 between Danish and Swedish animals, 0.46 between Danish and Finnish animals, and 0.86 between Swedish and Finnish animals.

Results from the combined reference data of the 3 populations (Table 5) show an average improvement in mean reliability of 0.03 for the Danish animals and 0.07 for the Swedish and Finnish animals, respectively, compared with using the national reference populations. A similar pattern was found when looking only at the youngest animals, but the reliabilities were, on average, slightly lower. The improvement was present across all traits, except for fertility in the Danish population. Comparison of the Swedish-Finnish reference data (Table 5) to the combined reference data shows that including Danish reference data only led to a very small change in mean reliability of DGV for Swedish and Finnish animals. For the Swedish animals, a 0.01 decrease occurred in mean reliability, and for the individual traits the predictions were either equally good or better in the full combined reference set. For the Finnish animals, a 0.01 increase occurred in mean reliability, and for the individual traits some differences were found between reliabilities of DGV from the 2 combined reference populations. No clear pattern was found in the changes of reliability of Finnish DGV when including or excluding Danish animals in the reference, likely because of the low genetic similarity between the 2 populations.

The results in this study support previous studies showing an increase in the reliability of genomic prediction when using combined reference populations (Hayes et al., 2009b; Su et al., 2009). On the other hand, adding Danish reference data to Swedish–Finnish reference data did not give a clear improvement in genomic prediction of Swedish and Finish animals. Two possible nonexclusive reasons exist: (1) a relatively lower genetic similarity between the Danish population and the other 2, compared with the genetic similarity between the Swedish and Finish populations and (2) only a small number of animals (778) were added.

In this study, published EBV of reference animals were used as response variable for genomic prediction, with focus on reliability of DGV. Garrick et al. (2009) point out problems with using EBV as phenotypes in genomic predictions. Estimated breeding values are already regressed, which could cause DGV of high-ranking animals to be underestimated and DGV of low-ranking animals to be overestimated. However, Su et al. (2010) pointed out "The advantage of using EBV is that they can be obtained directly from routine genetic evaluations. In addition, they contain little random error, which greatly reduces the prediction error variance. This could be important in situations where the number of genotyped animals in the reference is small." Results by Guo et al. (2010) showed that using EBV as response variables led to reliabilities of DGV slightly

Table 4. Reliabilities of genomic selection for the national studies and cross predictions for Danish (DK), Swedish (SWE), and Finnish (FIN) animals[1]

	Reference: DK			Reference: SWE			Reference: FIN		
Item	DK[2]	SWE	FIN	DK	SWE[3]	FIN	DK	SWE	FIN
Size of reference	929	778	778	1,395	1,551	1,395	1,562	1,562	1,562
Size of test	929	1,395	1,562	778	1,551	1,562	778	1,395	1,562
Milk yield	0.23	0.02	0.01	0.02	0.14	0.17	0.02	0.13	0.20
Fat yield	0.27	0.02	0.02	0.00	0.12	0.20	0.01	0.18	0.26
Protein yield	0.17	0.01	0.01	0.00	0.11	0.15	0.00	0.12	0.15
Yield index	0.17	0.01	0.01	0.01	0.11	0.16	0.00	0.14	0.17
Udder health	0.21	0.02	0.00	0.03	0.19	0.17	0.02	0.13	0.19
Fertility	0.24	0.09	0.03	0.03	0.18	0.14	0.03	0.16	0.19
Other diseases	0.43	0.04	0.00	0.08	0.28	0.12	0.01	0.10	0.25
Longevity	0.14	0.02	0.01	0.01	0.07	0.09	0.01	0.06	0.15
Direct calving	0.25	0.05	0.03	0.05	0.18	0.20	0.02	0.18	0.32
Maternal calving	0.26	0.01	0.05	0.03	0.04	0.09	0.04	0.04	0.16
Body	0.32	0.07	0.01	0.11	0.30	0.28	0.02	0.21	0.26
Feet and legs	0.25	0.13	0.03	0.13	0.32	0.13	0.05	0.20	0.10
Udder conformation	0.22	0.10	0.04	0.10	0.21	0.19	0.02	0.14	0.20
Milking ability	0.12	0.01	0.00	0.03	0.19	0.17	0.01	0.17	0.25
Temperament	0.25	0.11	0.04	0.11	0.31	0.12	0.04	0.20	0.20
Nordic total merit index	0.17	0.01	0.00	0.00	0.16	0.13	0.01	0.08	0.10
Growth	0.22	0.17	0.06	0.03	0.31	0.22	0.01	0.07	—
Mean (5-fold)	0.23	0.05	0.02	0.05	0.19	0.16	0.02	0.14	0.19
Mean (young)	0.19	—	—	—	0.19	—	—	—	0.18

[1]Mean reliability is also given using only the youngest cross-validation group. Standard deviation of reliabilities between the 5 cross-validation test data sets ranged between 0.05 and 0.1.

[2]Reference data including some Swedish bulls.

[3]Reference data including some Danish bulls

higher than those using daughter yield deviations as response variable in a simulation study, but underestimated variation of DGV because EBV is a regressed variable. On the other hand, previous studies showed an inflation of genomic prediction using deregressed proof as a response variable (Aguilar et al., 2010; Lund et al., 2010).

Table 6 shows the within-year regression coefficients of EBV on DGV (i.e., the annual means have been subtracted before doing the regression). In general, the regression coefficients show deviation from 1. For the Swedish animals, regression coefficients, however, have a larger-than-general deviation from 1, for both maternal calving and longevity, and for the Finnish animals, a large deviation is observed for feet and legs. These traits, however, have a very low reliability for all choices of reference populations in the mentioned test populations. Regression coefficients for protein yield, yield index, and the Nordic total merit index also seem to deviate significantly from 1 in all scenarios. Regression of EBV on DGV is, however, not well defined because it relies on the reliability of EBV and independence between prediction errors for EBV and DGV. When regressing true breeding value on DGV predicted using EBV as response variable, Guo et al. (2010) showed regression coefficients significantly higher than 1 in a simulation study. Therefore, regression of EBV on DGV might be not an appropriate measure of bias for genomic prediction.

The observed reliabilities in this study were lower than those reported from Holstein data. Hayes et al. (2009a) and Su et al. (2010) reported reliabilities between 0.20 and 0.70 for Holstein data. Comparison of results from Tables 4 and 5 with the reliabilities of parent averages given in Table 1 shows that genomic prediction only offers an advantage compared with the parent averages from a sire model for young bulls. Two possible explanations follow. (1) Differences in how the individual countries measure the traits, so the EBV explains different biological characteristics in different countries, thus making it more difficult to pick up the effects of QTL when combining the populations. The EBV are, however, predicted in a joint Nordic model, and experience from combining Nordic Holstein populations does not indicate that this should be a problem. Furthermore, mean reliabilities in Swedish and Finnish Red are lower than reported for the Danish Jersey, where a mean reliability of 0.30 across 5 traits has been observed (Thomasen et al., 2010), even though the size of the reference populations was smaller. (2) Jersey and Holstein results are obtained from populations with a more homogenous genetic composition than the Nordic Red. Many sires from different populations and breeds have been introduced in the Nordic Red cattle popula-

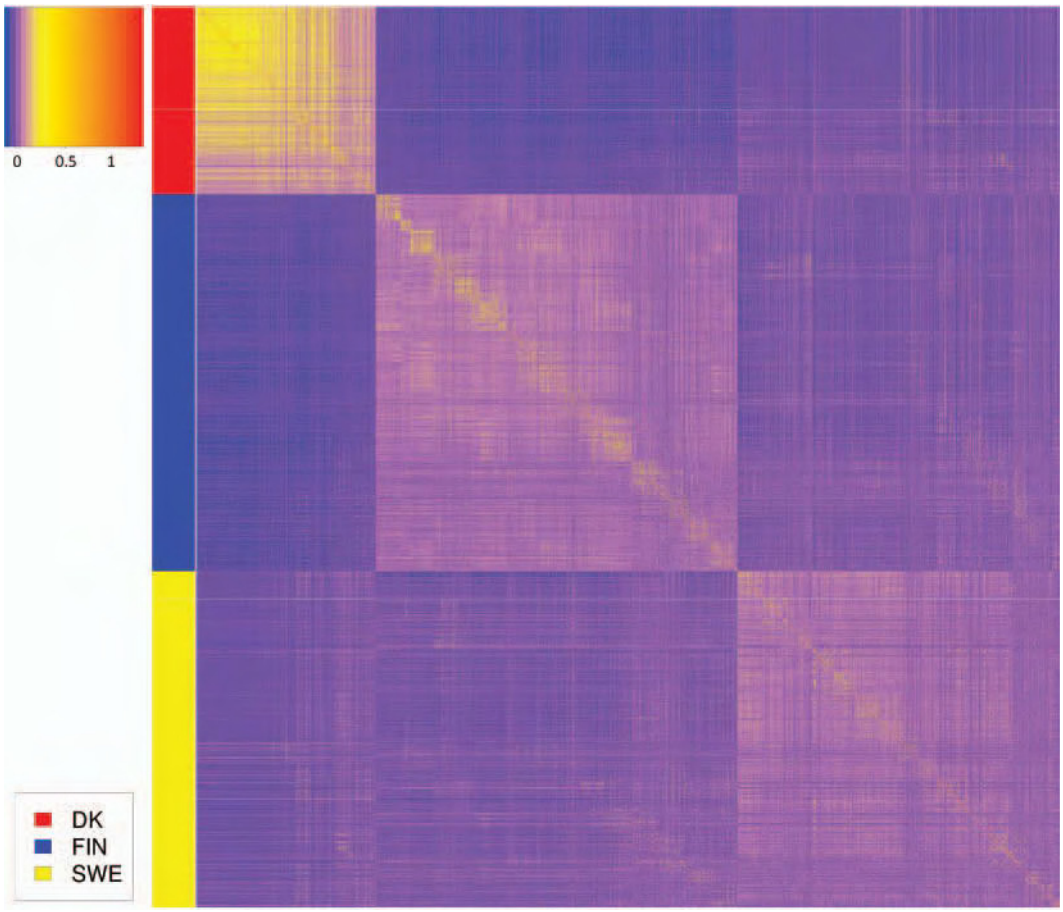

Figure 1. Genomic relationship matrix (**G**) for Danish (DK), Finnish (FIN), and Swedish (SWE) animals, sorted by country and year of birth. **G** is scaled to be analogous to the numerator relationship matrix (VanRaden, 2008).

tions and this gives a relatively higher heterozygosity, which might cause differences in LD patterns. An investigation of LD showed that mean squared correlations between markers in the interval 0 to 100kb are 0.25, 0.26, 0.21, 0.19, 0.20, and 0.19 for Danish Jersey, Nordic Holstein, Finnish Red, Danish Red, Swedish Red, and combined Red respectively (J. R. Thomasen, Viking Genetics International, Randers, Denmark, personal communication). The lower extent of LD in the Nordic Red populations could be the main reason for the lower reliabilities of DGV, compared with the Danish Jersey and Nordic Holstein populations. More detail is given by Rius-Vilarrasa et al. (2011).

It should be noted that the reported reliability of DGV is conservative in this study. The reliability was measured as the squared correlation between DGV and conventional EBV for the bulls in the test data, which were preselected based on parent average. The correlation between DGV and EBV for the selected animals would be lower than that for a random sample. In a previous study by Uimari and Mäntysaari (1993), it was derived that 10% selection based on parent average will decrease the correlation between parent average and daughter-based EBV by half (e.g., from 0.62 to 0.31). VanRaden et al. (2009) suggested to correct for selection by adding the difference between expected (from model) and observed (from validation) reliabilities of parent average EBV to the observed reliability of genomic prediction. However, this is only obtainable if the expected reliability of parent average EBV is unbiased.

Further studies are needed to exploit the large benefits of genomic selection for the Nordic Red. It has been argued that markers might not explain the whole additive genetic variance; therefore, a model including polygenic effects could increase the reliability of genomic prediction. However, Rius-Vilarrasa et al. (2010) reported that the improvement of genomic prediction by including a polygenic effect in the model for the Nordic red is small. Another possibility to increase the reliability of the genomic predictions is to use a blending procedure, where information from the DGV

Table 5. Reliabilities of genomic selection for combined Swedish-Finnish (SWE-FIN) reference and combined Danish-Swedish-Finnish (DK-SWE-FIN) reference data[1]

	Reference: SWE-FIN			Reference: DK-SWE-FIN			
Item	SWE	FIN	Both	DK	SWE	FIN	All
Size of reference	2,986	2,986	2,986	3,735	3,735	3,735	3,735
Size of test	1,411	1,575	2,986	778	1,395	1,562	3,735
Milk yield	0.21	0.29	0.26	0.25	0.23	0.25	0.27
Fat yield	0.26	0.37	0.32	0.31	0.30	0.31	0.34
Protein yield	0.20	0.25	0.23	0.18	0.21	0.18	0.22
Yield index	0.22	0.27	0.25	0.18	0.23	0.18	0.23
Udder health	0.25	0.29	0.28	0.23	0.25	0.23	0.28
Fertility	0.25	0.25	0.27	0.21	0.27	0.21	0.26
Other diseases	0.30	0.31	0.34	0.47	0.34	0.47	0.43
Longevity	0.11	0.20	0.16	0.15	0.10	0.15	0.15
Direct calving	0.25	0.37	0.32	0.28	0.27	0.28	0.32
Maternal calving	0.09	0.20	0.23	0.29	0.09	0.29	0.26
Body	0.38	0.14	0.41	0.38	0.41	0.38	0.55
Feet and legs	0.34	0.28	0.24	0.30	0.37	0.30	0.31
Udder conformation	0.26	0.31	0.29	0.27	0.29	0.27	0.38
Milking ability	0.25	0.26	0.29	0.18	0.29	0.18	0.29
Temperament	0.33	0.18	0.29	0.33	0.35	0.33	0.33
Nordic total merit index	0.20	0.14	0.21	0.19	0.18	0.19	0.19
Growth	0.36	0.17	0.36	0.29	0.32	0.29	0.32
Mean (5-fold)	0.25	0.27	0.28	0.26	0.26	0.26	0.30
Mean (young)	0.26	0.25	0.26	0.23	0.27	0.26	0.28

[1]Reliabilities are estimated for Danish (DK), Swedish (SWE), Finnish (FIN), Swedish-Finnish (both), and Danish-Swedish-Finnish (all) animals. Mean reliability is also given using only the youngest cross-validation group. The standard deviation of reliabilities between the 5 cross-validation test data sets ranged between 0.02 and 0.07.

is blended with the proofs from the traditional BLUP models. However, results from a separate study on blending for the Nordic Red breeds only showed an increase in reliability of about 1.5 percent for young animals (results not shown). One promising approach is to use the new high-density SNP chip. With a denser marker set, stronger LD would exist between markers and the QTL. This might allow the markers to explain more of the genetic variation, thus increasing the reliability (Harris et al., 2008), and would greatly benefit

Table 6. Within-year regression coefficients of EBV on direct genomic values (DGV) for Danish (DK), Swedish (SWE), and Finish (FIN) test animals[1]

	Ref: DK	Ref: SWE	Ref: FIN	Ref: SWE-FIN	Ref: SWE-FIN	Ref: SWE-FIN	Ref: all	Ref: all	Ref: all	Ref: all
Item	DK	SWE	FIN	SWE	FIN	SWE-FIN	DK	SWE	FIN	All
Milk yield	0.73	0.84	0.79	0.79	0.88	0.84	0.75	0.80	0.91	0.83
Fat yield	0.80	0.92	0.89	0.87	0.95	0.92	0.89	0.89	0.97	0.91
Protein yield	0.60	0.72	0.61	0.73	0.74	0.72	0.67	0.73	0.75	0.70
Yield index	0.61	0.75	0.65	0.75	0.76	0.75	0.70	0.75	0.77	0.73
Udder health	1.17	0.96	0.82	0.96	0.94	0.96	1.17	0.92	0.95	0.98
Fertility	0.96	0.91	0.83	0.90	0.89	0.91	0.86	0.90	0.84	0.88
Other diseases	1.09	0.90	0.81	0.89	0.84	0.90	1.04	0.92	0.81	0.95
Longevity	0.95	0.82	0.83	0.65	0.88	0.82	1.01	0.60	0.92	0.84
Direct calving	1.13	0.99	0.97	0.92	1.04	0.99	1.19	0.93	1.01	1.02
Maternal calving	1.01	0.85	0.88	0.67	0.89	0.85	1.12	0.62	0.86	0.91
Body	0.99	1.03	0.93	1.08	0.97	1.03	0.99	1.04	0.94	1.00
Feet and legs	1.06	0.80	0.57	1.07	0.55	0.80	1.04	1.06	0.55	0.87
Udder conformation	1.13	0.84	0.71	0.82	0.78	0.84	1.01	0.83	0.84	0.93
Milking ability	0.95	0.98	1.01	0.91	1.03	0.98	0.92	0.91	1.00	0.96
Temperament	0.97	0.89	0.88	0.93	0.86	0.89	0.85	0.89	0.86	0.88
Nordic total merit index	0.72	0.51	0.52	0.67	0.59	0.67	0.82	0.68	0.66	0.70
Growth	1.03	0.95	0.84	0.97	0.60	0.95	1.04	0.93	0.56	0.98
Mean	0.93	0.86	0.80	0.86	0.83	0.87	0.95	0.84	0.83	0.89

[1]DGV are predicted using either national (DK, SWE, or FIN), Swedish-Finish (SWE-FIN), or combined (all) reference (Ref) animals.

less homogeneous populations such as the Nordic Red. Another approach is to increase reference population size by international cooperation.

CONCLUSIONS

The aim of this study was to investigate the effects of including information from different, but genetically related, populations on the reliability of genomic selection. The reliability of genomic prediction for Nordic Red dairy cattle was improved by combining reference data from different Nordic Red populations. The benefit of combining reference populations was larger for the Swedish and Finnish populations, which have stronger genetic links, and less for the Danish population, which has fewer genetic links with the other two.

ACKNOWLEDGMENTS

This work was performed in the project Genomic Selection—From function to efficient utilization in cattle breeding (grant no. 3412-08-02253), founded by the Danish Directorate for Food, Fisheries and Agri Business (Copenhagen, Denmark); Viking Genetics (Randers, Denmark); Nordic Cattle Genetic Evaluation; Aarhus University (Tjele, Denmark); and the Swedish Farmers' Foundation for Agricultural Research (Stockholm, Sweden).

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.

Danish Cattle Federation. 2006. Principles of Danish cattle breeding. 8th ed. The Danish Agricultural Advisory Centre, Århus N, Denmark. Accessed June 24, 2011. http://www.landbrugsinfo.dk/Kvaeg/Avl/Sider/principles.pdf?List=%7B054f3326-7be6-421d-9e9a-d95e2403ae0c%7D&download=true.

Garrick, D. J., J. F. Taylor, and R. L. Fernando. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41:55.

Goddard, M. E., and B. J. Hayes. 2009. Mapping genes for complex traits in domestic animals and their use in breeding programmes. Nat. Rev. Genet. 10:381–391.

Guo, G., M. S. Lund, Y. Zhang, and G. Su. 2010. Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. J. Anim. Breed. Genet. 127:423–432.

Harris, B. L., D. L. Johnson, and R. J. Spelman. 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Pages 325–330 in Proc. Interbull Meeting, Niagara Falls, NY. Interbull, Uppsala, Sweden.

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

Hayes, B. J., P. J. Bowman, A. J. Chamberlain, K. Verbyla, and M. E. Goddard. 2009b. Accuracy of genomic breeding values in a multibreed dairy cattle population. Genet. Sel. Evol. 41:51.

Illumina Inc. 2005. Illumina GenCall Data Analysis Software: GenCall software algorithms for clustering, calling, and scoring genotypes. Illumina Pub. No. 370–2004–009. Illumina Inc., San Diego, CA. Accessed June 24, 2011. http://www.illumina.com/Documents/products/technotes/technote_gencall_data_analysis_software.pdf.

Luan, T., J. A. Woolliams, S. Lien, M. Kent, M. Svendsen, and T. H. E. Meuwissen. 2009. The accuracy of genomic selection in Norwegian Red cattle assessed by cross-validation. Genetics 183:1119–1126.

Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. Liu, R. Rents, C. Schrooten, M. Seefried, and G. Su. 2010. Improving genomic prediction by EuroGenomics collaboration. Article no. 880 in Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e. V., Giessen, Germany.

Matukumalli, L. K., C. T. Lawley, R. D. Schnabel, J. F. Taylor, M. F. Allan, M. P. Heaton, J. O'Connell, S. S. Moore, T. P. L. Smith, T. S. Sonstegard, and C. P. Van Tassell. 2009. Development and characterization of a high density SNP genotyping assay for cattle. PLoS ONE 4:e5350. doi:10.1371/journal.pone.0005350.

Meuwissen, T. H. E., B. J. Hayes, and M. E. Goddard. 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829.

Rius-Vilarrasa, E., R. F. Brøndum, I. Strandén, B. Guldbrandtsen, E. Strandberg, M. S. Lund, and W. F. Fikse. 2010. Bayesian models for predicting genomic breeding values in a Swedish-Finnish Red breed cattle population. Article no. 441 in Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e. V., Giessen, Germany.

Rius-Vilarrasa, E., T. Iso-Touru, I. Strandén, N. Schulman, B. Guldbrandtsen, E. Strandberg, M. S. Lund, J. Vilkki, and W. F. Fikse. 2011. Characterization of linkage disequilibrium in a Danish, Swedish and Finnish Red breed cattle population. Abstracts of the 62nd European Federation for Animal Science (EAAP) Annual Meeting, Stavanger, Norway. Wageningen Academic Publishers, Wageningen, the Netherlands.

Su, G., B. Guldbrandtsen, V. R. Gregersen, and M. S. Lund. 2010. Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. J. Dairy Sci. 93:1175–1183.

Su, G., G. Guo, and M. S. Lund. 2009. Accuracy of genomic evaluations depends on distance to the reference data. Page 296 in Abstracts of the 60th European Federation for Animal Science (EAAP) Annual Meeting, Barcelona, Spain. Wageningen Academic Publishers, Wageningen, the Netherlands.

Team Avlsværdivurdering. 2009. Årsstatistik Avl—2008/2009. Dansk Kvæg, Skejby, Denmark.

Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brøndum, and M. S. Lund. 2010. Reliabilities of genomic estimated breeding values in Danish Jersey. Article 839 in Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. Gesellschaft für Tierzuchtwissenschaften e. V., Giessen, Germany.

Uimari, P., and E. Mäntysaari. 1993. Repeatability and bias of estimated breeding values for dairy bulls and bull dams calculated from animal model evaluations. Anim. Prod. 57:175–182.

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

VanRaden, P. M., C. P. Van Tassell, G. R. Wiggans, T. S. Sonstegard, R. D. Schnabel, J. F. Taylor, and F. S. Schenkel. 2009. Reliability of genomic predictions for North American Holstein bulls. J. Dairy Sci. 92:16–24.

Villumsen, T. M., L. Janss, and M. S. Lund. 2009. The importance of haplotype length and heritability using genomic selection in dairy cattle. J. Anim. Breed. Genet. 126:3–13.