The admixed population structure in Danish Jersey dairy cattle challenges accurate genomic predictions

J. R. Thomasen,*† 2 A. C. Sørensen,† G. Su,† P. Madsen,† M. S. Lund,† and B. Gulbrandtsen†

*VikingGenetics, Ebeltoftvej 16, 8860 Assentoft, Denmark; and †Center for Quantitative Genetics and Genomics, Faculty of Science and Technology, Aarhus University, DK-8830, Tjele, Denmark

ABSTRACT: The main purpose of this study was to evaluate whether the population structure in Danish Jersey (DJ) known from the history of the breed also is reflected in its genomic structure. This is done by comparing the linkage disequilibrium and persistence of phase for subgroups of Jersey animals with high proportions of Danish (DNK) or United States (USJ) origin. Furthermore, it is investigated whether a model explicitly incorporating breed origin of animals, inferred either through the known pedigree or from SNP marker data, leads to improved genomic predictions compared with a model ignoring breed origin. The study of the population structure incorporated 1,730 genotyped Jersey animals. In total 39,542 SNP markers were included in the analysis. The 1,079 genotyped bulls with de-regressed proof for udder health were used in the analysis for the predictions of the genomic breeding values. A range of random regressions models that included the breed origin were analyzed and compared with a basic genomic model that assumes a homogeneous breed structure. The main finding in this study is that the importation of germplasm from the USJ population is readily reflected in the genomes of modern DJ animals. First, linkage disequilibrium in the group of admixed DJ animals is lower compared with the groups of the original DNK and USJ animals. Second, persistence of linkage disequilibrium phase is not conserved for longer marker distances between animals with mainly Danish or United States origin. Third, the STRUCTURE analysis could retrieve genomic-based breed proportions in alignment to the pedigree-based breed proportions. However, including this population structure in a random regression prediction model did not clearly improve the reliabilities of the genomic predictions compared with a basic genomic model.

Key words: cattle, genomic predictions, population structure

INTRODUCTION

Genomic predictions are more accurate than traditional pedigree-based evaluations (Meuwissen et al., 2001) when the evaluation of the animal does not include progeny information. The important factors influencing reliabilities of genomic predictions (Hayes et al., 2009) are the number of animals with phenotypes in the genotyped reference population, the reliability of the phenotypes that are used to predict the SNP marker effects, and the level of linkage disequilibrium (LD) between SNP markers and QTL. The size of the reference population and phenotypic recording strategies can be changed through changes in the breeding strategy. In contrast, the level of LD at a given distance between loci is a function of the population history (Hayes et al., 2009).

Danish Jersey (DJ) is an example of an admixed breed consisting of animals with diverse origin. It includes animals with different breed proportions of original Danish Jersey (DNK) and United States Jersey (USJ). The 2 populations share common ancestry even though they have been separated for almost a century. However, due to the long separation,
phase associations between marker and QTL alleles may differ depending on the origin of the chromosome segments, thus reducing LD across the 2 substructures in DJ. These effects may explain why predicted reliabilities in DJ are less than expected given the number and quality of phenotypes in the reference population (Thomasen et al., 2012).

The main objective of this study is to investigate whether the historic pedigree population structure in DJ is reflected in its genomic structure. This is investigated by LD measures and persistency of marker phase within and between subgroups of animals with high and low proportions of DNK (vs. USJ) in their pedigree. In addition, it is investigated whether explicitly accounting for the population structure by genomewide grouping of animals improves the genomic predictions. A random regression linear model is used to model the population structure in the genomic predictions and compared with a traditional pedigree model and a genomic model that assumes a homogeneous breed structure.

**MATERIALS AND METHODS**

Routine collections by trained staff of semen and blood samples were used for the genotypings.

**Genetic Data**

All animals included in this study were genotyped using the Illumina Bovine SNP50 BeadChip (Matukumalli et al., 2009). The SNP typing was done in part at the Department of Molecular Biology and Genetics, Aarhus University, and in part at Genoscan A/S (Tjele, Denmark). The genotypic data was edited by individual typing for each animal. For animals the requirements was a call rate above 95% except for 10 animals (0.6% of total), which were accepted with call rates of at least 85%. Marker loci were accepted if they had a call rate of at least 95%. Loci with a minor allele frequency less than 1% were excluded. Loci without a map position in the UMD 3.1 assembly (Zimin et al., 2009) were discarded. Animals with an average GenCall (GC) score of less than 0.65 were discarded. Individual marker genotypings with a GC score of less than 0.6 were discarded.

After marker data quality checking 39,542 informative SNP markers were available for in total 1,730 Jersey animals born between 1983 and 2010. Due to missing DNA samples only 18 bulls born in the years 1983 to 1987 were included. The dataset consisted of 1,480 genotyped Jersey sires and 250 genotyped cows. The fastPHASE software (Scheet and Stephens, 2006) was used to impute sporadic missing genotypes and to phase the SNP markers for the Jersey animals.

**Breed Origin of Animals**

For the study of the LD structures, 2 subgroups were performed based on the breed proportions traced by the pedigree. Animals with at least 75% Danish origin in their pedigree were assigned to the DNK group consisting of 171 animals. Animals with at least 75% USJ origin in pedigree were assigned to the USJ group consisting of 131 animals. In this study 52 original USJ sires and 27 original DNK sires are included. Breed proportions of Canadian Jersey were included in USJ. Due to import of semen from a few New Zealand bulls in the 1960s many of the genotyped Jersey animals have a small part of their pedigree originated from the New Zealand Jersey population. The breed proportions with this origin were not used in this study.

The average proportion of original DNK varies from 0.25 for bulls born in 1985 up to 0.75 for the bulls born in 1988. The proportion of DNK has stabilized around 0.60 for bulls born after 2002. The SD in breed proportion of DNK has been reduced over the years showing that the animals are continuously admixed.

**Estimation of Linkage Disequilibrium and Persistence of Phase**

The $r^2$ was used as the measure for the LD (Lewontin, 1964) as it seems to be the most robust measure for the LD (Qanbari et al., 2010). The $r^2$ was calculated as the squared correlations between markers that were grouped into bins by lengths of 7 kb.

Persistence of marker phase was calculated as the Pearson correlation $\rho$ of $r$ across the groups DJ, USJ, and DNK, in which $r$ is the correlation between markers within bin and subpopulation of a defined marker interval. The persistence of marker phase across the groups was calculated within bins of 25 kb in marker distances (de Roos et al., 2008). The larger marker distance for calculations of the $r^2$ was used to reduce stochastic sampling variance of the calculated $\rho$. The minimum number that was included in the calculation of $\rho$ was 10,598 in 1 bin for the persistence between USJ and DNK.

**Historic Population Size**

The present effective population size ($N_e$) is a result of its historic population evolution. Traditionally estimated $N_e$ is based on the pedigree (Nomura, 1996). However, with the availability of dense markers both present and historic $N_e$ can be estimated from LD information.
The historic population size \((N_T)\), \(T\) generations ago, for the DJ population was inferred from the calculated \(r^2\) over a certain interval between loci. The expected \(r^2\) (Sved, 1971) can for small values of marker intervals \(c\) (in morgans) and in the absence of mutations be calculated as \(r^2 = 1/(4N_e c + 1)\), in which \(N_e\) is the effective population size and \(c\) is the linkage map distance in morgans. By reformulating this formula the expected value of \(N_T\) can be calculated as \(E(N_T) \approx (1/\sigma^2 - 1)/4c\), in which \(T = 1/2c\) (Hayes et al., 2003). Using the calculated \(r^2\) for each bin of 7 kb marker distances \(E(N_T)\) were calculated within each bin.

**Measures of Breed Origin**

The traditional genomic prediction model assumes that QTL effects are in perfect LD with the surrounding SNP. However as hypothesized in this paper the LD may be reduced due to the admixture structure in DJ and hence decrease the reliability of the genomic predictions if this substructure is not accounted for in the model.

We used 2 measures of breed origin in the genomic prediction model to account explicitly for the population structure: 1) the pedigree-based breed proportion and 2) a marker-based breed proportion. Marker information may give a more reliable estimate of the population structure if the pedigree data is not complete (Flury et al., 2010). Therefore breed proportions were also inferred from the marker information using the software package STRUCTURE (Pritchard et al., 2000). The linkage model that uses the positions of the individual markers was chosen to infer the breed proportions (Falush et al., 2003). The model accounts for both 1) LD due to mixture of the subpopulations that arise from variation in pedigree and leads to correlation among markers even if they are unlinked and 2) LD due to admixture that can be recognized from each of the ancestral populations. However, the linkage model ignores the common background LD that persists across the subpopulations. Therefore this model is best suited to use loosely linked markers (Falush et al., 2003). A marker spacing of around 1 cM was therefore used. This was achieved by choosing 1 out of every 100 SNP markers without any other restrictions. In total, 412 evenly spaced markers were chosen across all 29 autosomes. In the STRUCTURE analysis 2 subpopulations were assumed \((k = 2)\). Based on our knowledge of the history in the breed, using \(k = 2\) is a strong prior. We also tested \(k = 3\) because a small proportion of the present DJ animals originate from the New Zealand Jersey population. However, it was not possible to infer this subpopulation as a third breed in the STRUCTURE analysis. No prior pedigree information about the breed origin was used to infer the marker-based breed proportion. For burn-in, the first 10,000 iterations of the Gibbs sampler were discarded. After burn-in, the Markov chain Monte Carlo (MCMC) sampler was iterated 20,000 times. This is the standard setting in the program, which was found to be sufficient for the present dataset. In a study by Frkonja et al. (2012) including only 10,000 MCMC were performed.

**Dependent Variable**

De-regressed proofs (DRP) for udder health were used as response variable for the prediction of genomic breeding values. Official EBV from January 2011 was used for the calculations of DRP for udder health (Danish Cattle Federation, 2012). This trait was chosen because there is a relatively large difference in genetic level for this trait between the USJ and DNK groups (Danish Cattle Federation, 2012), indicating that there have been different breeding goals in the 2 populations. The EBV are corrected for dominance effects due to total heterosis and recombination loss but include differences in genetic levels between original USJ and DNK animals (Negussie et al., 2010). In total, 1,079 genotyped animals had records for DRP. Calculations of the DRP followed the procedures described by Strandén and Mäntysaari (2010). The DRP for udder health varied between 60.10 and 126.7 with a variance of 118.2. Reliabilities ranged from 47 to 99%.

**Genomic Predictions**

The 1,079 bulls with DRP for udder health were split into a reference population consisting of 879 bulls and a test population consisting of the 200 youngest bulls. Udder health was chosen because it has low heritability with a high economic value in the breeding goal. Use of marker information for increasing reliability of prediction is therefore particularly beneficial. Variance components were estimated using data from the reference population using AI REML. Next, genomic breeding values were predicted using phenotypes and genotypes from the reference population and genotypes from the test population. The DMU software package (Madsen and Jensen, 2010) was used for the estimation of the variance components and for prediction of the genomic breeding values.

The statistical model used for the prediction of genomic breeding values explicit accounting for the population structure was

\[
y = 1\mu + w\beta + Za_0 + Wa_i + e, \tag{1}
\]

in which \(y\) is the vector of DRP for udder health, \(1\) is a vector of ones, \(\mu\) is the intercept, \(w\) is a vector with element \(w_i\) as the proportion of Danish origin for animal \(i\) from either pedigree or STRUCTURE, \(\beta\) is the fixed regression of DRP on breed proportion, the \(i\)th element
in $a_0$ is the random animal intercept for animal $i$, the $i$th element in $a_1$ is the random regression on breed proportion for animal $i$, $Z$ is an incidence matrix associating $a_0$ with $y$, $W$ is a matrix with 1 nonzero element in each row presenting the proportion of Danish origin for corresponding animal, and $e$ is the vector of residuals. It is assumed that $e \sim N(0,D \sigma^2_e)$ in which $D$ is a diagonal matrix having the diagonal elements $d_i = (1 - \text{REL}_{\text{DRP}})/\text{REL}_{\text{DRP}}$, in which $\text{REL}_{\text{DRP}}$ is the reliability of DRP for animal $i$ and $\sigma^2_e$ is the residual variance. We used this procedure to correct for heterogeneous residual variance due to different reliabilities of DRP.

The random animal effects $a_0$ and $a_1$ are assumed to follow the distribution:

$$
\begin{pmatrix}
a_0 \\
a_1
\end{pmatrix} \sim N\left(0, \begin{pmatrix}
\sigma_{a0}^2 & \sigma_{a0a1} \\
\sigma_{a0a1} & \sigma_{a1}^2
\end{pmatrix} \otimes G^* \right)
$$

in which $G^*$ is the weighted genomic and pedigree relationship matrix (Gao et al., 2012). A weight of 0.05 on the pedigree relationship matrix is used. Different weights of the pedigree relationship matrix were tested but were not found to influence the predictive ability.

The direct genomic breeding value (DGV) for animal $i$ is calculated as

$$
\text{DGV}_i = \beta w_i + a_0 i + w_i a_1 i. \quad [2]
$$

Ignoring the population structure Eq. [1] can be reduced to

$$
y = 1 \mu + Za_0 + e. \quad [3]
$$

Formula [3] represents the basic genomic model that assumes a homogeneous breed structure. The DGV for animal $i$ is now reduced to

$$
\text{DGV}_i = a_0 i. \quad [4]
$$

Finally, a prediction model equivalent to [3] but using the pedigree-based relationship was used to predict breeding values to evaluate the advantage of genomic prediction.

**Reliability of Direct Genomic Breeding Value**

Reliability of genomic predictions has in several studies been assessed as the validation reliability, $r^2_v$ (Su et al., 2010; Lund et al., 2011; Thomasen et al., 2012). The validation reliability of the DGV was calculating as the squared correlation between DGV and DRP divided by the mean reliability of DRP for the bulls in the test data set, which was in line with the definition of reliability of estimated breeding value because

$$
r^2_v = \frac{\text{Cov}(\text{DGV}, \text{DRP})}{\text{Cov}(\text{DGV}, a) / \sigma_{\text{DGV}}^2} \approx \frac{\text{Cov}(\text{DGV}, a) / \sigma_{\text{DGV}}^2}{\text{Cov}(\text{DGV}, \text{DRP}) / \text{REL}_{\text{DRP}}},
$$

in which $a$ was the true breeding value (Su et al., 2012). The predictive ability for the random regression model [1] was compared with the basic genomic model [3].

**RESULTS**

**The Level of Linkage Disequilibrium**

The level of $r^2$ in the admixed group of animals (DJ), as shown in Fig. 1, is marginally less than in the 2 subgroups (DNK and USJ). For a marker spacing of 100 kb the level of LD is 0.02 greater within the 2 subgroups compared with the admixed group including all animals. This difference in $r^2$ is constant up to a distance of 700 kb. The level of $r^2$ decreases with increasing marker distances and decrease to a level of 0.08 for all animals included in the calculations (DJ).

**Persistence of Phase**

The greatest levels of phase persistence were observed between the DJ and DNK groups, which for the marker spacing interval 100 to 200 kb has a level of 0.95 (Fig. 2). Phase persistence between USJ and DJ was slightly less, decreasing to a level of 0.91 at marker space of 100 to 200 kb. Least persistence is seen between the DNK and the USJ group with a value of 0.78 for the marker spacing interval 100 to 200 kb.

**Historic Population Size**

Historic effective population sizes estimated from LD data up to 20 generations in the past are shown in Fig. 3. The effective population size shows a minimum
4 to 6 generations ago with a level of 135. This is the time period when the import of USJ germplasm started to play a major role in the breeding program for DJ (P. G. Larson, VikingGenetics, Randers, Denmark, personal communication). As a consequence a small increase in the effective population size was seen 3 to 4 generations in the past. From generation 20 and down to generation 6 in the past the estimated effective population size decreases approximately linearly from 190 to 135.

Population Structure

A comparison of the subpopulation structure either inferred from STRUCTURE or traced by the pedigree is shown in Fig. 4. Overall the comparison between the 2 measures shows a correlation of 0.81 with a regression of 1.01. This correlation is obtained without attaching any predefined breed origin to the animals defined as either USJ or DNK from the pedigree. For a given proportion of DNK origin expected from the pedigree the figure visualize the variation in proportion of DNK origin inferred from the markers. Animals that are estimated to be purebred DNK (proportion = 1) based on the pedigree were seen to have inferred breed proportion varying between 0.71 and 1, except for 1 animal. This indicates that some of purebred DNK are in fact admixed with USJ or they still carry haplotypes inherited from the shared founder population. The STRUCTURE analysis with the SNP markers was able to identify the admixture in the genome. The animals that are estimated to be purebred USJ (proportion = 0) have breed proportion varying between 0.05 and 0.20. The USJ animals are on average 7 yr younger and therefore more distant from the common founder population.

Genomic Predictions

Table 1 shows the reliabilities of the genomic predictions for udder health using different prediction models. The letters referring to the specific models are given in Table 1 The greatest reliability (33.7%) for the test bulls is obtained for the genomic model (model f) that uses the pedigree-derived breed proportion as covariate for both fixed and random regression. In this model the covariance between the random animal effects \((a_0\) and \(a_1\)) was restricted to 0. This model made only slightly (0.08%) better predictions than the basic genomic model (model d) without information on population structure. In general, there are only small differences (0.9%) in predictive ability between all the genomic prediction models. The basic animal model (model h) that uses the traditional pedigree-based relationship matrix as covariance structure has a predictive ability of 24.2%. Thus, the model using marker information (model d) increases predictive ability by 9.5%. The 3 models (e, f, and g) that use pedigree derived breed proportions as population structure information on average perform 0.30% better than marker-derived breed proportions. However, the correlation between the pedigree and marker derived breed proportions were relatively high (0.81), indicating that marker inferred breed proportions is consistent with the breed proportions assigned to pedigree.

DISCUSSION

Population Structure

The main finding in this study is that the importation of germplasm from the USJ population is readily reflected in the genomes of modern DJ animals. First, \(r^2\) in the group of admixed DJ animals are lower compared with the groups of the original DNK and USJ animals. Second, persistence of LD phase is not conserved for longer marker distances (>40 kb) between DNK and USJ animals. As we only have a relative few purebred original USJ and DNK animals included in the present study, the differences in \(r^2\) and persistence of phase between purebred USJ and purebred DNK animals will be even greater than found
in this study. Third, the STRUCTURE analysis could retrieve genomic-based breed proportions in alignment to the pedigree-based breed proportions.

In general the pattern of the $r^2$ for marker space differences up to 100 kb observed in the present study follows the results obtained in other genetic structure analysis of the cattle breeds (Gautier et al., 2007; McKay et al., 2007; de Roos et al., 2008; Pryce et al., 2010). For larger SNP marker spacing (up to 1 Mb) the $r^2$ in the DJ population are greater than $r^2$ found in the Australian (Pryce et al., 2010) and New Zealand (de Roos et al., 2008) Jersey populations. These differences might reflect that the recent historic population development in the DJ has been more closed and selection more intensive, leading to a smaller effective population size.

The pattern of persistence of phase and levels of LD reflect the history of the USJ and DNK populations as we know that both subpopulations originate from the Channel Island of Jersey, more than 20 generations in the past. DJ originates from a single import of 5,200 animals from the Channel Island of Jersey around year 1900. The animals that founded the USJ population were imported to North America around year 1860. Until the 1960s DS was developed a closed breed without further import of animals or semen. From this period the import of USJ semen has started (P. G. Larson, VikingGenetics, Randers, Denmark, personal communication).

The STRUCTURE analysis gave inferred breed proportions in alignment with pedigree-based breed proportions and a regression coefficient between the 2 measures close to 1. The correlation of 0.81 is a bit lower than obtained by Frkonja et al. (2012), using the same number of markers (0.93). However, Frkonja et al. (2012) predefined the breed for the purebred animals, which increases the correlation. The 2 analyzed breeds in their study (Simmental and Red Holstein) have been separated for many more generations than the 2 Jersey populations in the present study have. The analysis also supported our prior expectation that DJ consists of 2 major subpopulations. We used the linkage model in STRUCTURE, which accounts for LD due to the mixture of populations and LD due to admixture (Falush et al., 2003). Background LD, however, is not accounted for. Therefore, we chose a set of loosely linked markers as suggested by Falush et al. (2003). Increasing the number of markers will increase the amount of background LD in the data, which can lead to misleading results. The relatively large variation in inferred breed proportion obtained in STRUCTURE for a given value of breed proportion from pedigree may reflect that the separation of the 2 subpopulations has been fairly short. Therefore, some genomic segments cannot be associated with a particular subpopulation.

### Effective Population Size

The estimated current effective population size in DJ of 135 based on the markers is in alignment with the marker-based estimate found in the New Zealand Jersey population (de Ross et al., 2008), which also was estimated to 135. Sørensen et al. (2005) calculated an effective population size of 53 in DJ based on the rate of inbreeding calculated from the pedigree. The estimate using pedigree information is highly dependent on mating decisions as it measures the correlation of uniting gametes. The marker-based population size is in contrast a function of associations of alleles in gametes within and between animals. The increase in effective population size 3 to 5 generations in the past indicates that the import of USJ semen has introduced new combinations of marker alleles in the DJ population. This might be explained by a differentiation in the breeding goals between the 2 subpopulations and genetic drift within the populations.

<table>
<thead>
<tr>
<th>Model</th>
<th>Population structure information</th>
<th>Covariance (%) of predictions for test bulls</th>
</tr>
</thead>
<tbody>
<tr>
<td>(a) $y = I\mu + W\beta + Z_{a0} + W_{a1} + e$</td>
<td>Markers</td>
<td>Genomic</td>
</tr>
<tr>
<td>(b) $y = I\mu + W\beta + Z_{a0} + W_{a1} + e^2$</td>
<td>Markers</td>
<td>Genomic</td>
</tr>
<tr>
<td>(c) $y = I\mu + W\beta + Z_{a0} + e$</td>
<td>Markers</td>
<td>Genomic</td>
</tr>
<tr>
<td>(d) $y = I\mu + Z_{a0} + e$</td>
<td>None</td>
<td>Genomic</td>
</tr>
<tr>
<td>(e) $y = I\mu + W\beta + Z_{a0} + W_{a1} + e$</td>
<td>Pedigree</td>
<td>Genomic</td>
</tr>
<tr>
<td>(f) $y = I\mu + W\beta + Z_{a0} + W_{a1} + e^2$</td>
<td>Pedigree</td>
<td>Genomic</td>
</tr>
<tr>
<td>(g) $y = I\mu + W\beta + Z_{a0} + e$</td>
<td>Pedigree</td>
<td>Genomic</td>
</tr>
<tr>
<td>(h) $y = I\mu + Z_{a0} + e$</td>
<td>None</td>
<td>Pedigree</td>
</tr>
</tbody>
</table>

1Population structure information is obtained either from the markers or from the traditional pedigree. The information for the covariance structure of animals in the model is derived from either genomic or traditional pedigree information.

2In model [b] and [f], $\sigma_{a0}^2$ was set to 0.
Genomic prediction in an admixed population

Using Population Structure in Genomic Predictions

The genomic prediction model that has been validated in the DJ population (Thomassen et al., 2012) assumes a homogenous population structure in DJ. This assumption implies that LD between markers and QTL persists for all animals. However, we have shown in this study that persistence of marker phase is not unique within the DJ breed for marker distances above 10 kb when information about the subpopulation structure is studied. We therefore hypothesized that differences in marker allele effects exist due to difference in origin of alleles (DNK vs. USJ) and that use of the genetic structure either based on the information from the markers or the pedigree can improve the prediction of the genomic breeding values.

Several studies exploit merging of reference populations, as the size of the reference population has been shown to affect the reliabilities of the genomic predictions considerably (Goddard and Hayes, 2009). Merging the largest European Holstein reference populations improved the reliabilities considerably (Lund et al., 2011) whereas combination of the Nordic Red populations only resulted in minor improvement of the reliabilities (Brøndum et al., 2011). A further extension of the Nordic Red reference population with the Norwegian Red even slightly decreased the reliabilities for some fertility traits (Heringstad et al., 2011). The Nordic Red is a group of 3 cattle populations, which again includes contributions from several breeds. Persistence of phase varied from 0.80 and up to 0.94 for marker intervals up to 500 kb (Rius-Vilarrasa et al., 2011). This indication of slightly different associations might be the explanation for the small improvement when the Nordic Red populations are combined into 1 reference population.

Prediction Models Including Population Structure

The overall conclusion from comparing different genomic prediction models is that including population structure in a random regression prediction model did not clearly improve the reliabilities of the genomic predictions compared with a basic genomic model. There can be several explanations for the lack of improvement. First, the population structure might already be modeled well by the genomic relationship matrix. A simulation study for admixed and crossbred populations (Toosi et al., 2010) showed that the genomic selection prediction equations performed well as long as all purebreds that contributed to the population were included in the reference set, even without origin of breed having been taken explicitly into account. Second, the population structure information was based on average marker information measure across the entire genome. This might be a too naive measure to model the differences in marker allele effects as admixed animals consist of a mosaic of haplotypes originating from purebreds DNK and USJ animals. Therefore, more detailed random regression models should be set up, which allows the breed origin of markers within an individual to vary over the genome. For instance a specific marker allele might be linked to a positive udder health effect when it originates from a DNK animal but linked to a negative effect when it has a USJ origin. However, due to high overall positive correlations between SNP effects in the 2 breeds of origin, it was in the present study not possible to separate these effects, using average genomewide population structure information.

Genomic Predictions across Populations

Danish Jerseys are a relative small population with only 1,000 bulls with phenotypic information. As reliabilities increases with the size of the reference populations, use of phenotypic information from other populations may be a strong tool to increase the reliabilities for the genomic predictions. Genomic predictions across populations were evaluated in a simulation study by de Roos et al. (2009). The conclusion from their study was that reliabilities of genomic predictions could be improved if reference data from all populations were included in the reference set. However, it requires that LD phase persists across populations. If the persistence between populations was low, reliabilities might even be reduced (de Roos et al., 2009). Investigation of persistence of marker phases between Jersey and the other Nordic dairy cattle populations showed that the persistence of marker phases is low. Results from that study are not shown in this paper. Adding reference data from either Holstein or Nordic Red is therefore not expected to increase reliabilities. It will at least require a high density marker genotyping of all the Nordic reference populations and a more sophisticated modeling of population structure before the benefits from combination of the Nordic reference populations can be realized.

LITERATURE CITED


