Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds

B.C.D. Cuyabano, G. Su, G.J.M. Rosa, M.S. Lund, D. Gianola

Bootstrap study of genome-enabled prediction reliabilities using haplotype blocks across Nordic Red cattle breeds

B. C. D. Cuyabano,[*1] **G. Su,**[*1] **G. J. M. Rosa,†** **M. S. Lund,*** **and D. Gianola†**
*Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, 8830 Tjele, Denmark
†Department of Animal Sciences, University of Wisconsin, Madison 53706

ABSTRACT

This study compared the accuracy of genome-enabled prediction models using individual single nucleotide polymorphisms (SNP) or haplotype blocks as covariates when using either a single breed or a combined population of Nordic Red cattle. The main objective was to compare predictions of breeding values of complex traits using a combined training population with haplotype blocks, with predictions using a single breed as training population and individual SNP as predictors. To compare the prediction reliabilities, bootstrap samples were taken from the test data set. With the bootstrapped samples of prediction reliabilities, we built and graphed confidence ellipses to allow comparisons. Finally, measures of statistical distances were used to calculate the gain in predictive ability. Our analyses are innovative in the context of assessment of predictive models, allowing a better understanding of prediction reliabilities and providing a statistical basis to effectively calibrate whether one prediction scenario is indeed more accurate than another. An ANOVA indicated that use of haplotype blocks produced significant gains mainly when Bayesian mixture models were used but not when Bayesian BLUP was fitted to the data. Furthermore, when haplotype blocks were used to train prediction models in a combined Nordic Red cattle population, we obtained up to a statistically significant 5.5% average gain in prediction accuracy, over predictions using individual SNP and training the model with a single breed.
Key words: bootstrap analysis, haplotype block, multi-breed genomic prediction, Nordic Red cattle

INTRODUCTION

Genome-enabled prediction methods based on markers, such as SNP, have been widely explored in animal breeding since such methods were introduced (Meuwissen et al., 2001). Haplotype blocks (haploblocks) have been extensively studied in human genetics (Curtis et al., 2001; Chapman et al., 2003; Curtis, 2007), and were explored in animal breeding in the early days of genomic prediction. With advances in SNP technology, a trend emerged toward using SNP genotypes in genome-enabled prediction. More recently, there is renewed interest in the use of haploblocks in prediction of livestock traits for various reasons, one of them being use of information across breeds. Studies have suggested that use of haploblocks can lead to a higher prediction accuracy than use of SNP markers (Villumsen et al., 2009; Boichard et al., 2012; Cuyabano et al., 2014).

A basic assumption in genome-enabled selection is that each QTL is in linkage disequilibrium (**LD**) with some surrounding markers, given that marker panels are sufficiently dense. Genome-enabled prediction uses this assumption as a basis for defining models that estimate the effects of markers on a phenotype of interest. A main potential advantage of haploblocks (i.e., a group of nearby SNP) over individual SNP markers is that each haploblock may be in higher LD with the causative mutations than would be any individual SNP.

One important feature of haploblocks over individual SNP for prediction is that alleles within a haploblock may capture more variation. As a simple example to illustrate this, take 2 biallelic loci A and B of a phased genotype. In the standard regression on SNP genotypes, phenotype y is fitted by equation $\hat{y} = \hat{\mu}_{SNP} + \hat{g}_A A_1 + \hat{g}_B B_1$, where A_1 is 1 if locus A has allele 1 and 0 otherwise, and the same applies to B_1; $\hat{\mu}_{SNP}$ is the mean phenotype when alleles A_2 and B_2 are observed; and $\hat{g}$ are the estimated effects of the respective loci. In a regression on haploblock genotypes, phenotype y is fitted as $\hat{y} = \hat{\mu}_{hap} + \hat{g}_1 A_1 B_1 + \hat{g}_2 A_1 B_2 + \hat{g}_3 A_2 B_1$, where $A_i B_j$ is 1 if locus A has allele i and locus B has allele j ($i,j = 1,2$) observed, and zero otherwise. When dealing with haploblocks built based on LD, the grouping of many SNP into a haploblock can reduce the number of variables needed to perform genomic prediction (Cuyabano et

Table 1. Example of estimated haplotype effects by regressing on SNP or haplotype block (haploblock) genotypes, where $\hat{g}$ are the estimated SNP/haploblock effects obtained by the models

	Haplotype effect on phenotype y	
Haplotype	Individual SNP	Haploblocks
A_1B_1	$\hat{g}_A + \hat{g}_B$	$\hat{g}_1$
A_1B_2	$\hat{g}_A$	$\hat{g}_2$
A_2B_1	$\hat{g}_B$	$\hat{g}_3$
A_2B_2	0	0

al., 2014). Furthermore, LD-based haploblocks do not have a fixed number of SNP per haploblock. Table 1 indicates the estimated effects for regressions on both SNP and haploblock genotypes of each haplotype allele on phenotype y. Using the regression on SNP genotypes, the effect of A_1B_1 is the sum of the effects of A_1B_2 and A_2B_1, whereas haplotypes may have effects that are not linear (e.g., interactions). This may influence accuracy of predictions.

Our hypothesis is, therefore, that haploblocks may improve prediction of economically important traits. To test this, genome-enabled predictions obtained using haploblocks were evaluated and compared with those obtained by using a regression on SNP genotypes. Furthermore, use of haploblocks in an across-breed scenario for prediction may increase predictive ability, mainly due to an increase in the variation captured by haploblock alleles. When working with a single breed, we may not observe some haploblock alleles that occur only in another breed. Hence, the use of a combined population allows us to estimate the effects of haploblock alleles that might not be observed in a single breed analysis.

It is known that when different populations have a common origin, combining them for the training of prediction models can improve prediction reliability (Lund et al., 2011). The stronger the genetic ties between the populations, the greater the benefits (Su et al., 2009). With regard to combined data of different breeds, an increase in prediction reliabilities was reported for Danish, Finnish, and Swedish Red cattle populations when training was based on the combined data instead of using individual breeds (Brøndum et al., 2011). Hence, our work also aimed to assess the potential benefit from using a combined population for training models to obtain genomic predictions of breeding values.

This work presents results of predictions for 5 traits in dairy cattle production: fertility, mastitis, and yields of protein, fat, and milk. We compared the predictive ability of models using regressions on SNP or on haploblock genotypes, and using either a single breed or a combined population of Nordic Red cattle (including

Danish, Finnish, and Swedish Red) for model training. We were interested in comparing results when using a Bayesian BLUP model or a Bayesian mixture model to infer allelic effects, to verify whether the model would have an influence on the prediction results. We also wished to measure uncertainty associated with results, which is crucial for proper statistical interpretation. Thus, bootstrap sampling was used in the test data set to generate samples of prediction reliabilities. Samples were analyzed with multivariate statistical methods in an innovative comparison of prediction reliabilities.

MATERIALS AND METHODS

Marker and Phenotypic Data

The marker and phenotypic data used was from a sample of 4,403 animals from the Nordic Red cattle population. The original marker data were obtained from a 54K (~54,000 markers) SNP chip and then imputed to 777K data using Beagle (Browning and Browning, 2009; Brøndum et al., 2012; Su et al., 2012). The imputed data were then edited by removing markers in complete LD with adjacent markers and with a minor allele frequency (MAF) <0.01 (Su et al., 2012). After editing, a final marker data set with 442,267 SNP for the 4,403 animals was obtained. This population comprised 3 sub-breeds: Danish Red (**DR**), Swedish Red (**SRB**), and Finnish Ayrshire (**FAY**). The data set was split into training and test data sets, using as cut-off the birth date of bulls of October 1, 2001, resulting in a training population of 3,423 animals (663 DR, 1,051 SRB, and 1,709 FAY) and a test population of 980 animals (186 DR, 306 SRB, and 488 FAY).

The phenotypic values used to obtain genome-enabled predictions of fertility, mastitis, protein, fat and milk yield were deregressed proofs (**DRP**) of the bulls, derived from the EBV and their effective daughter contributions (Jairath et al., 1998; Schaeffer, 2001; Garrick et al., 2009).

Animal Ethics

The phenotypic data were collected from routine records of dairy cattle farms. Genotyped animals used in this work were progeny-tested bulls, and the semen samples for genotyping were obtained from routine bull semen collection. Therefore, no ethical approval was necessary.

Haplotype Blocks

The haploblocks were built based on LD measured as D' (Gabriel et al., 2002; Cuyabano et al., 2014). The

use of LD for defining haploblocks allows the latter to differ in number of SNP per haploblock, instead of arbitrarily defining haploblocks with a fixed length. This nonrandom setting of where in the genome a haploblock begins and ends increases the LD between SNP within a haploblock and reduces the total number of explanatory variables to be included in the prediction models.

Based on a previous study (Cuyabano et al., 2014), we determined that a haploblock would be defined by a group of sequential SNP satisfying $D' > 0.45$ between every pair of SNP in this group. A total of 82,824 "multiallelic" haploblocks were obtained from the high-density marker data. Because each haploblock may have more than 2 possible "alleles," a total of 318,113 haploblock variables were obtained for performing genome-enabled prediction. Haploblocks included from 1 to 62 SNP, and 90% of them containing ≤10 SNP. The number of "alleles" per haploblock varied from 2 to 18, and 91% of the haploblocks had no more than 7 different "alleles."

Genomic Prediction

Predictions of the 5 traits of interest were performed on scenarios differing on the covariates (SNP or haploblocks), the training data set (combined Red cattle population or individual Red cattle breed) and the statistical model employed (Bayesian BLUP or Bayesian mixture model), producing a total of 8 different scenarios when combining these 3 factors. The focus was on the differences between the combinations of the covariates genotypes and training populations. Therefore, the analyses were always separated by the statistical method; that is, Bayesian BLUP or mixture.

The prediction models were defined by the standard equation $\mathbf{y} = \mathbf{1}\mu + \mathbf{Mg} + \mathbf{Za} + \boldsymbol{\varepsilon}$, where $\mathbf{y}$ represens the vector containing the DRP of training bulls, μ is a general mean; $\mathbf{M}$ is the SNP/haploblock matrix with elements coded as 0, 1, or 2, and in the case of haploblocks, for each haploblock with k different alleles, there are $k - 1$ columns coded as 0, 1, or 2, indicating the counts of the haploblock's alleles; $\mathbf{g}$ is the vector of additive genetic effects; $\mathbf{Z}$ is the incidence matrix linking $\mathbf{a}$ to $\mathbf{y}$; $\mathbf{a}$ is the vector of residual polygenic effects; and $\boldsymbol{\varepsilon}$ is the vector of the random residuals of the model. It was assumed for both Bayesian BLUP and Bayesian mixture models that $\mathbf{a} \sim N\left(0, \mathbf{A}\sigma_a^2\right)$, where $\mathbf{A}$ was the genetic relationship matrix constructed according to the pedigree, and $\boldsymbol{\varepsilon} \sim N\left(0, \mathbf{D}\sigma_\varepsilon^2\right)$, where $\mathbf{D}$ is a diagonal matrix with $d_{ii} = \left(1 - r_{DRP_i}^2\right)\big/r_{DRP_i}^2$, and $r_{DRP_i}^2$ is the trait reliability of the ith animal (VanRaden, 2008; Garrick et al., 2009). For the Bayesian BLUP, we as-

sumed that $\mathbf{g} \sim N\left(0, \mathbf{I}\sigma_g^2\right)$ and, for the Bayesian mixture (George and McCulloch, 1993; Meuwissen, 2009; Gao et al., 2013), that $\mathbf{g} \sim \sum_{k=1}^{4} \pi_k N\left(0, \mathbf{I}\sigma_{\pi_k}^2\right)$. The mixing proportions were fixed as $\pi_1 = 0.889$, $\pi_2 = 0.1$, $\pi_3 = 0.01$, and $\pi_4 = 0.001$, and the variances were estimated under the constraint $\sigma_{\pi_1}^2 < \sigma_{\pi_2}^2 < \sigma_{\pi_3}^2 < \sigma_{\pi_4}^2$. The prior distribution for the mean μ was uniform within the $(-\infty, \infty)$ interval, and all variances were within $(0, \infty)$ and assumed to have independent uniform priors.

The statistical models were implemented using a Bayesian Markov chain Monte Carlo algorithm with the BayZ package (Janss, 2011), running a single chain with length 50,000, in which the first 20,000 cycles were taken as the burn-in of the chain.

Evaluation of Prediction Results with Bootstrapping

After running the prediction models, the analysis and comparisons of predictive ability were performed using bootstrap samples (Efron, 1979; Efron and Tibshirani, 1986) of the test animals. One hundred randomized bootstrap samples were generated, and each sample contained a random subset of two-thirds of the animals in the test data set. The number of individuals in the data set was decided mainly because of the relatively small test population of Danish Red. If too few individuals were used to compute the prediction reliabilities in each bootstrap sample, the variance would be large. Furthermore, using 100 bootstrap samples was assumed sufficient to evaluate the variability observed in the prediction reliabilities. A small test using 200 bootstrap samples did not produce a noticeable difference in variance of prediction reliability. The bth bootstrap sample and genomic estimated breeding value (**GEBV**) for the ith bull were represented as the pair $\left(DRP_i^b, GEBV_i^b\right)$ and the prediction reliability was calculated for each bootstrap sample. In this way, 100 observed prediction reliabilities were obtained for each prediction scenario.

The GEBV obtained from the prediction models were calculated as $GEBV_i = \sum_j m_{ij}\hat{g}_j + \hat{a}_i$, where $\hat{g}_j$ and $\hat{a}_i$ are the predictors of the additive genetic and residual polygenic effects, respectively. Reliabilities of the predictions of breeding values were obtained as the squared correlation between DRP and GEBV of individuals in the sample of the test population, divided by the average reliability of DRP, r_{DRP} (Garrick et al., 2009). Thus, for each bootstrap sample, prediction reliability was calculated as $r_b = Cor^2\left(\boldsymbol{DRP}^b, \boldsymbol{GEBV}^b\right)\big/r_{DRP}$. The performance of each model was assessed by analyz-

ing the estimated reliabilities of GEBV from the bootstrap samples.

Based on the reliabilities samples, confidence ellipses (Johnson and Wichern, 2007) for the mean reliabilities were calculated and graphed separately for the 2 statistical models (Bayesian BLUP and Bayesian mixture) to evaluate statistical differences of the prediction reliabilities obtained from different scenarios.

To construct the confidence ellipses, we took the observations $\left(\boldsymbol{R}_{i,SNP}, \boldsymbol{R}_{i,hap}\right)$, where $\boldsymbol{R}_{i,SNP}$ and $\boldsymbol{R}_{i,hap}$ were the bootstrap reliabilities (r_b) obtained from predictions using individual SNP or haploblocks, respectively. The index i was for the training data set ($i = c,s$) such that c indicated training on combined Nordic Red cattle population and s training on a single breed. Fisher's z-values (Fisher, 1915) were applied to the reliabilities,

$$z = \frac{1}{2} log\left(\frac{1 + \sqrt{r}}{1 - \sqrt{r}}\right). \quad [1]$$

This transformation of the reliabilities approximates a bivariate normal distribution of the pairs. Therefore, $\left(\mathbf{Z}_{i,SNP}, \mathbf{Z}_{i,hap}\right) \sim N\left(\boldsymbol{\mu}_i, \boldsymbol{\Sigma}_i\right)$, where $\boldsymbol{\mu}_i = \left[\boldsymbol{\mu}_{i,SNP}, \boldsymbol{\mu}_{i,hap}\right]$ and $\boldsymbol{\Sigma}_i = Cov\left(\mathbf{Z}_{i,SNP}, \mathbf{Z}_{i,hap}\right)$, where $\mathbf{Z}$ are vectors of the transformed values from $\boldsymbol{R}$. All analyses were performed on the transformed data. Subsequently, we computed the following statistic (T):

$$T_i^2 = n_b \left(\overline{\mathbf{Z}}_i - \boldsymbol{\mu}_i\right)' \mathbf{S}_i^{-1} \left(\overline{\mathbf{Z}}_i - \boldsymbol{\mu}_i\right), \quad [2]$$

where $n_b = 100$ is the number of bootstrap samples, $\overline{\mathbf{Z}}_i = \left(\overline{Z}_{i,SNP}, \overline{Z}_{i,hap}\right)'$ and $\mathbf{S}_i = \hat{Cov}\left(\mathbf{Z}_{i,SNP}, \mathbf{Z}_{i,hap}\right)$. Because of the normality assumption for the z-transformed prediction reliabilities,

$$\frac{\left(n_b - 2\right)}{2\left(n_b - 1\right)} T_i^2 \sim F_{2,(n_b-2)}. \quad [3]$$

Finally, the 95% confidence ellipsis for $\left(\mathbf{Z}_{i,SNP}, \mathbf{Z}_{i,hap}\right)$ was defined by the coordinates

$$\left[\overline{\mathbf{Z}}_i - q\mathbf{S}_i^{1/2}\sqrt{\frac{2\left(n_b - 1\right)}{n_b\left(n_b - 2\right)}} \; ; \; \overline{\mathbf{Z}}_i + q\mathbf{S}_i^{1/2}\sqrt{\frac{2\left(n_b - 1\right)}{n_b\left(n_b - 2\right)}}\right], \quad [4]$$

where $q > 0$ is such that $P\left(\frac{\left(n_b - 2\right)}{2\left(n_b - 1\right)} T_i^2 \le q^2\right) = 0.95$.

One of our objectives was to evaluate improvements in prediction reliability. The focus was to assess the gain in predictive ability obtained from changing the covariates from a regression on SNP genotypes and training in a single breed, to regressing on haploblock genotypes and training on the combined Red cattle population. Aiming to obtain reliable comparisons, instead of analyzing the difference between the mean prediction reliabilities or the mean Fisher z-transformed data $\left(\Delta R = \overline{R}_{c,hap} - \overline{R}_{s,SNP} \text{ and } \Delta Z = \overline{Z}_{c,hap} - \overline{Z}_{s,SNP}\right)$, both a gross comparison), we calculated the statistical distance (Mahalanobis, 1936) between the centers of the confidence ellipses, as follows:

$$\Delta M = \sqrt{\frac{\left(\overline{Z}_{c,hap} - \overline{Z}_{c,SNP}\right)^2}{Var\left(\overline{Z}_{c,hap} - Z_{c,SNP}\right)} + \frac{\left(\overline{Z}_{s,hap} - Z_{s,SNP}\right)^2}{Var\left(\overline{Z}_{s,hap} - Z_{s,SNP}\right)}}. \quad [5]$$

Clearly, ΔR and ΔZ do not address variability, whereas ΔM measures more accurately an improvement in predictive ability.

Finally, to assess the importance of the factors evaluated in the different prediction scenarios, ANOVA (Patterson and Thompson, 1971) was carried out based on the bootstrapped samples.

RESULTS

Table 2 presents the P-values of the ANOVA performed to evaluate factors relevant to variations of prediction reliabilities. The ANOVA was done separately for each trait and for the 3 test populations and 2 prediction models. For both prediction models, the training population had a significant effect on prediction reliabilities for all traits in all populations, with P-values near zero. The genotype covariates; that is, SNP or haploblocks, were significantly different mainly when Bayesian mixture models were used. In these models, with the exception of the prediction of fertility and milk yield for DR and the prediction of protein for FAY, the predictors (SNP or haploblock genotypes) were significant either as a main effect or as an interaction with the training population. In the Bayesian BLUP models, on the other hand, the effect of the predictors was not significant in more than half of the scenarios.

Figures 1, 2, 3, 4, and 5 present the 95% confidence ellipses defined by equation [4] for the bootstrap samples obtained from the predictions of fertility, mastitis, protein, fat, and milk yield, respectively. The ellipses allow clear visualization of the dispersion of the prediction reliabilities obtained in each scenario. Each figure presents the ellipses separated in different panels corresponding to the breed predicted (DR, FAY, and SRB) and the statistical model used (Bayesian mixture or Bayesian BLUP). In each panel, there are 2 confidence

Table 2. *P*-values of the ANOVA of z-transformed prediction reliability by trait (fertility, mastitis, protein, fat, and milk yield) model (Bayesian BLUP or Bayesian mixture), predictor (haploblocks or SNP), and training population (combined Red cattle or single breed)[1]

Trait	Test population[2]	Bayesian mixture			Bayesian BLUP		
		Predictors	Training population	Predictors × training	Predictors	Training population	Predictors × training
Fertility	DR	0.225	**0.000**	0.627	0.063	**0.000**	0.711
	FAY	**0.015**	**0.000**	0.647	0.878	**0.000**	0.051
	SRB	**0.000**	**0.000**	**0.000**	0.646	**0.000**	0.546
Mastitis	DR	**0.000**	**0.000**	**0.000**	0.372	**0.000**	0.144
	FAY	*0.502	**0.000**	**0.000**	*0.454	**0.000**	**0.004**
	SRB	**0.000**	**0.000**	0.078	0.524	**0.000**	0.221
Protein	DR	*0.082	**0.000**	**0.004**	0.112	**0.000**	0.488
	FAY	0.170	**0.000**	0.914	*0.557	**0.000**	**0.000**
	SRB	**0.002**	**0.000**	**0.001**	**0.008**	**0.000**	0.990
Fat	DR	*0.998	**0.000**	**0.000**	*0.531	**0.000**	**0.003**
	FAY	**0.000**	**0.000**	0.068	0.219	**0.000**	0.132
	SRB	*0.434	**0.000**	**0.008**	**0.048**	**0.000**	0.178
Milk yield	DR	0.185	**0.000**	0.225	0.535	**0.000**	0.467
	FAY	**0.000**	**0.000**	**0.000**	**0.021**	**0.000**	**0.004**
	SRB	**0.015**	**0.000**	0.735	*0.132	**0.000**	**0.029**

[1]*P*-values in bold are those for significant factors on prediction reliability.

[2]DR = Danish Red; FAY = Finnish Ayrshire; SRB = Swedish Red.

P-values of individual factor effects with $P > 0.05$ but significant interaction between the factors (predictors × training population).

ellipses, one for training the model on the single breed and another for training the model in the combined Nordic Red cattle population. This allows comparison of prediction reliabilities obtained when combining these 2 factors. When the confidence ellipses in each panel overlap, it means that there is no significant statistical difference between predictions. In most cases, the confidence ellipses overlapped. The exceptions are the prediction of mastitis for FAY, the prediction of protein for SRB, the prediction of fat for FAY (only using the Bayesian BLUP model) and for SRB, and the prediction of milk yield for SRB. When comparing the results in these cases to those obtained with the ANOVA, it is possible to identify the significance of the training population. The covariates were also significant in these cases, either as an individual effect, or by interaction with the training population.

Table 3 presents the difference in prediction reliability between genome-enabled predictions obtained using either SNP and training on a single breed or from models using haploblocks and training on the combined Nordic Red cattle population. The direct differences (ΔR and ΔZ) in mean prediction reliabilities and the z-transformed data were much larger than the standardized differences (ΔM). As explained in the Materials and Methods section, this is because ΔR and ΔZ do not take into account the bivariate distribution and variances of the data. The discussion of gains will be based on ΔM only. Because this measure is on the same scale for all scenarios and traits, it allows direct comparisons, indicating which traits and populations benefit more from the multi-breed training population

combined with use of haploblock genotypes. For prediction of mastitis in FAY, protein in SRB, fat in FAY (using the Bayesian BLUP model only) and in SRB, and milk yield in SRB, the confidence ellipses showed a significant increase in prediction reliability, ranging from 3.7 to 5.5%.

DISCUSSION

This work studied the use of haploblocks as covariates in genome-enabled prediction models, as well as the effect of training the models on a combined Nordic Red cattle population. Previous studies suggested that haploblocks could improve prediction accuracy (Villumsen et al., 2009; Boichard et al., 2012; Cuyabano et al., 2014). Also, use of combined populations for training prediction models had shown benefits in prediction accuracy (Brøndum et al., 2011), especially when genetic ties between populations are strong (Su et al., 2009).

Instead of simply comparing prediction reliabilities between different scenarios, we extracted bootstrap samples from the test populations. Then, we generated a sample of realized prediction reliabilities, for each scenario. This allowed us to use multivariate statistical methods in the analyses of results.

In studies that compare genome-enabled prediction methods, it is common to calculate prediction reliabilities employing the whole test data and simply calculate differences. However, from a statistical point of view, this is equivalent to carrying out an experiment using one observation (a single prediction reliability is obtained for each method). Bootstrapping offers a fea-
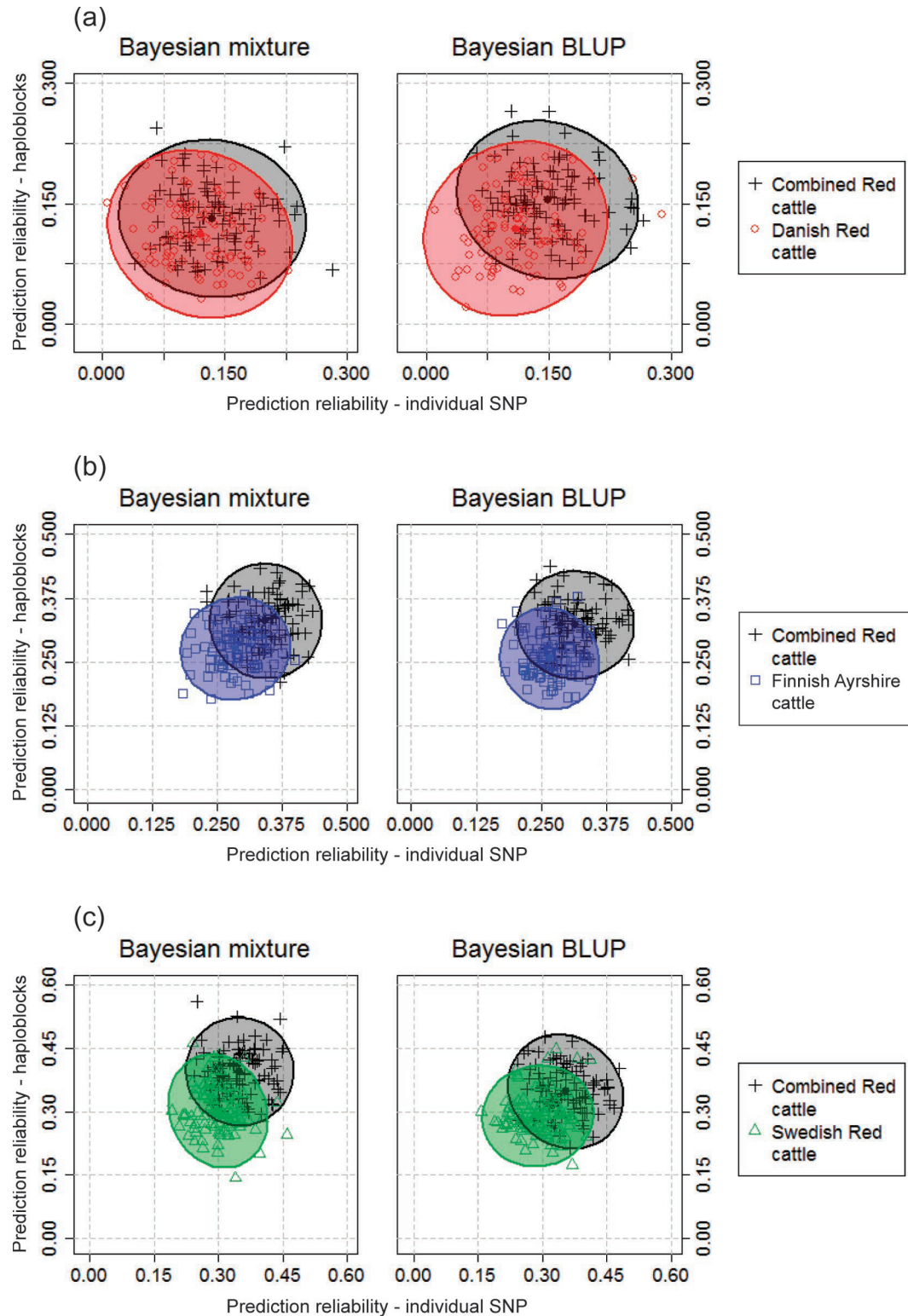
Figure 1. Fertility: The x-axis of the panels represents the mean prediction reliabilities of the bootstrap samples regressing on SNP genotypes. The y-axis represents the mean prediction reliabilities of the bootstrap samples regressing on haploblock genotypes. Scatter plot and confidence ellipses of predicted genomic EBV of (a) a Danish Red cattle population; (b) a Finnish Ayrshire cattle population; and (c) a Swedish Red cattle population obtained by bootstrapping the test data set. Color version available online.
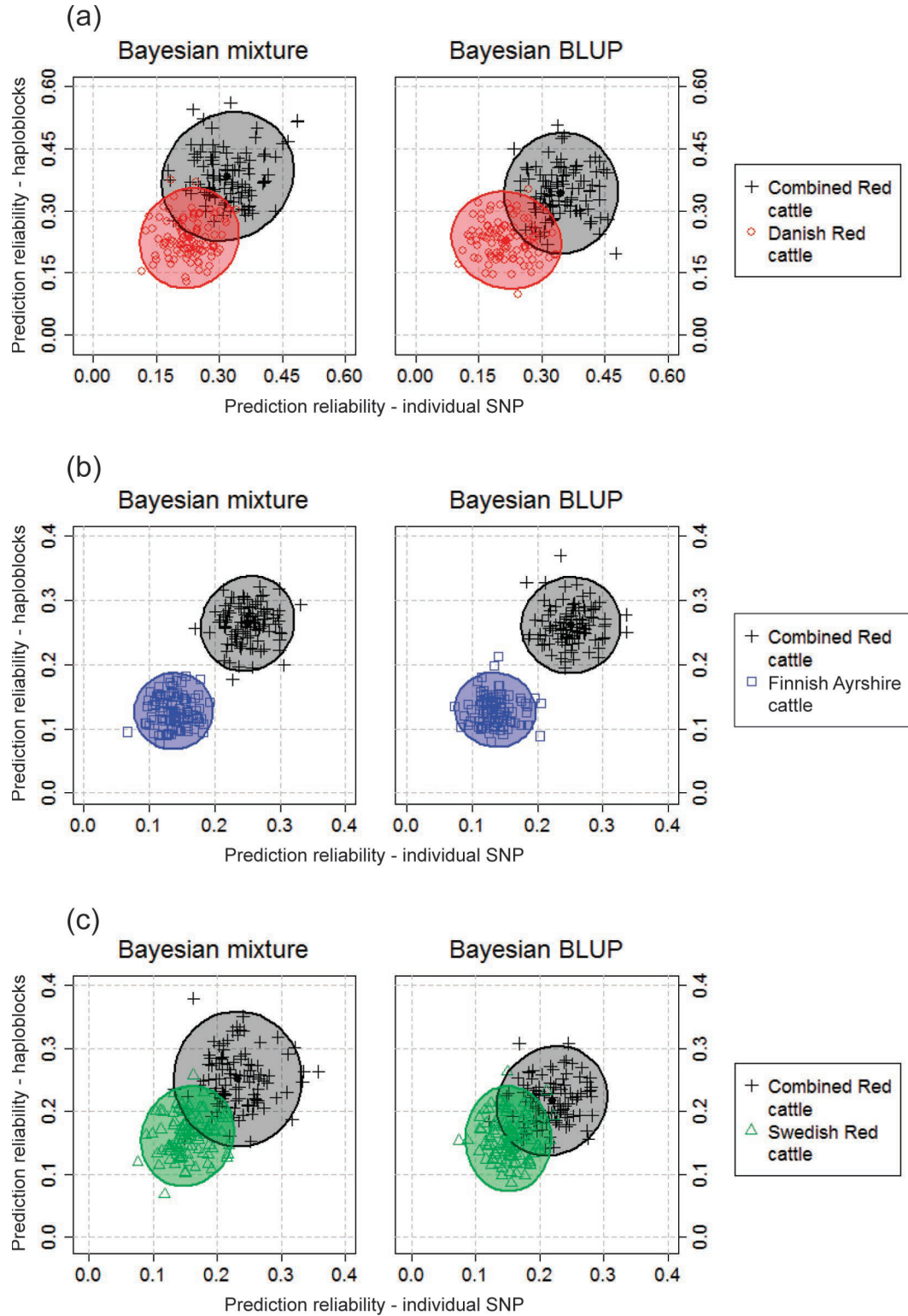
Figure 2. Mastitis: The x-axis of the panels represents the mean prediction reliabilities of the bootstrap samples regressing on SNP genotypes. The y-axis represents the mean prediction reliabilities of the bootstrap samples regressing on haploblock genotypes. Scatter plot and confidence ellipses of predicted genomic EBV of (a) a Danish Red cattle population; (b) a Finnish Ayrshire cattle population; and (c) a Swedish Red cattle population obtained by bootstrapping the test data set. Color version available online.
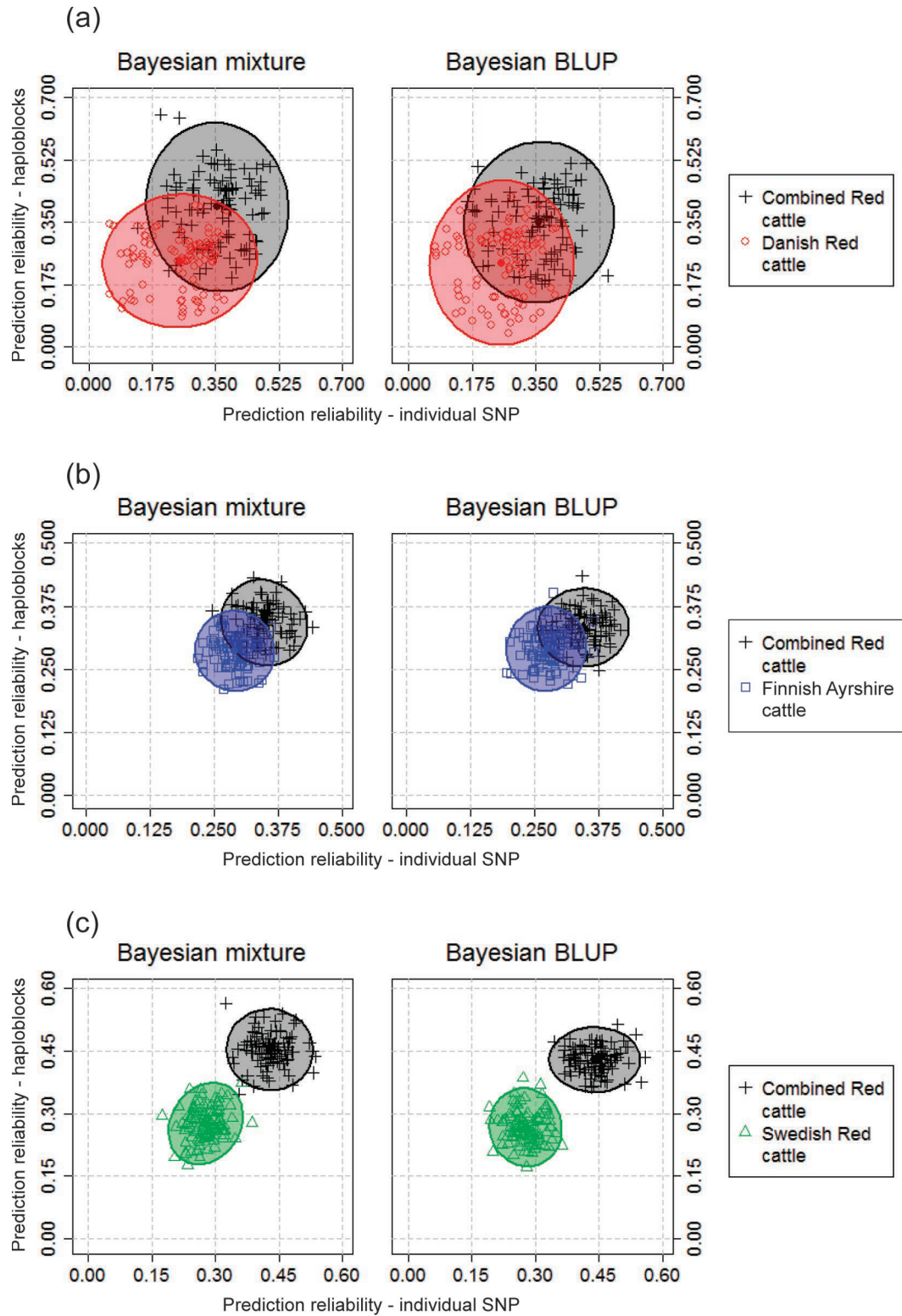
Figure 3. Protein: The x-axis of the panels represents the mean prediction reliabilities of the bootstrap samples regressing on SNP genotypes. The y-axis represents the mean prediction reliabilities of the bootstrap samples regressing on haploblock genotypes. Scatter plot and confidence ellipses of predicted genomic EBV of (a) a Danish Red cattle population; (b) a Finnish Ayrshire cattle population; and (c) a Swedish Red cattle population obtained by bootstrapping the test data set. Color version available online.

Figure 4. Fat: The x-axis of the panels represents the mean prediction reliabilities of the bootstrap samples regressing on SNP genotypes. The y-axis represents the mean prediction reliabilities of the bootstrap samples regressing on haploblock genotypes. Scatter plot and confidence ellipses of predicted genomic EBV of (a) a Danish Red cattle population; (b) a Finnish Ayrshire cattle population; and (c) a Swedish Red cattle population obtained by bootstrapping the test data set. Color version available online.
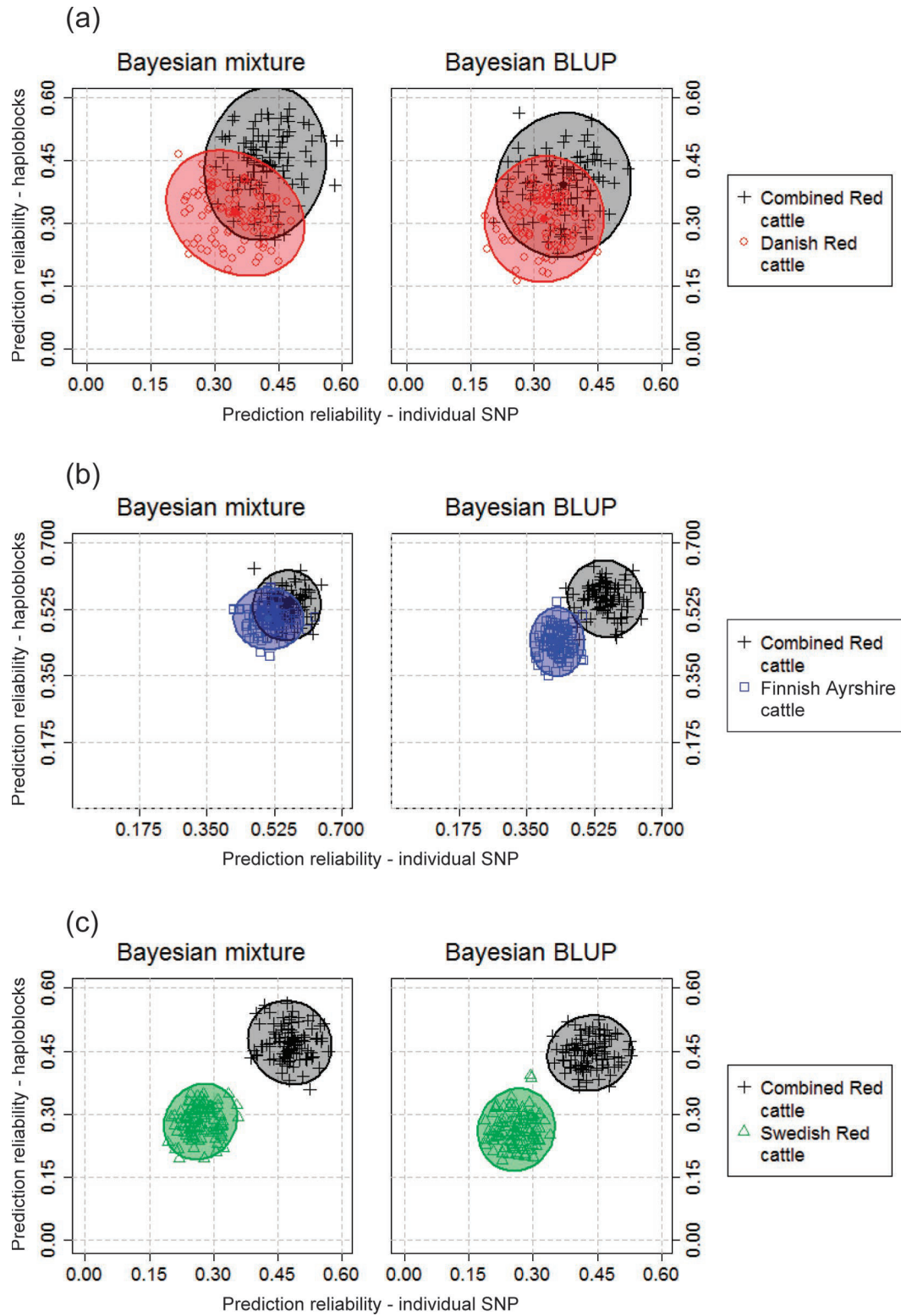
Figure 5. Milk yield: The x-axis of the panels represents the mean prediction reliabilities of the bootstrap samples regressing on SNP genotypes. The y-axis represents the mean prediction reliabilities of the bootstrap samples regressing on haploblock genotypes. Scatter plot and confidence ellipses of predicted genomic EBV of (a) a Danish Red cattle population; (b) a Finnish Ayrshire cattle population; and (c) a Swedish Red cattle population obtained by bootstrapping the test data set. Color version available online.
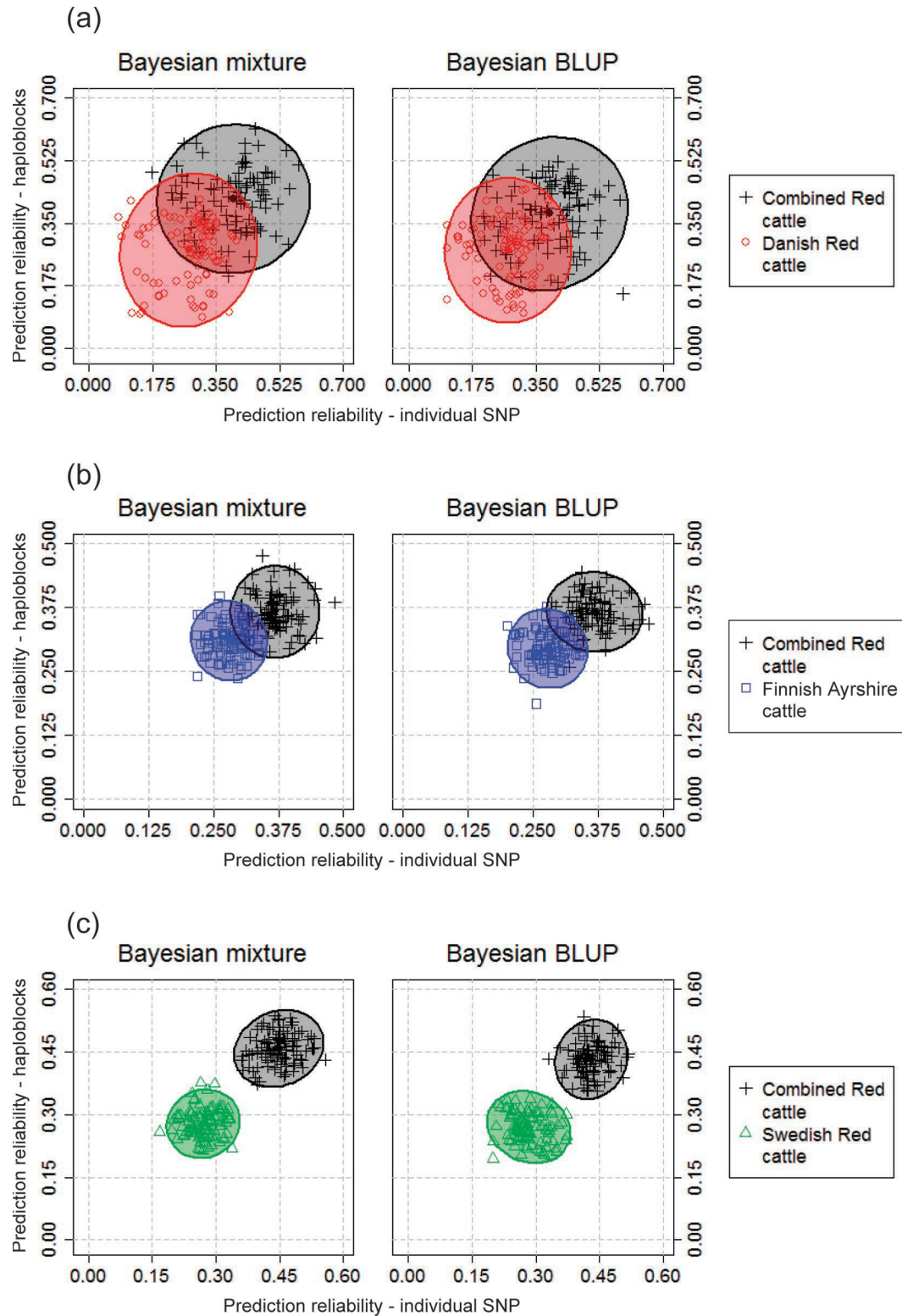
Table 3. Gain (%) obtained in mean prediction reliability of bootstrap genomic EBV, between genomic using haploblocks with a combined Red cattle population, and using SNP with a single reference population[1]

Trait	Test population[2]	Bayesian mixture			Bayesian BLUP		
		ΔR	ΔZ	ΔM	ΔR	ΔZ	ΔM
Fertility	DR	0.7	1.2	0.4	2.6	4.5	0.9
	FAY	2.3	4.1	1.3	2.9	5.3	1.3
	SRB	4.6	8.5	1.4	2.9	5.3	1.4
Mastitis	DR	7.7	13.9	2.2	6.5	11.6	2.3
	FAY	7.0	12.2	4.7	7.0	12.0	4.7
	SRB	5.4	9.5	2.3	3.7	6.4	1.8
Protein	DR	6.9	12.6	1.5	4.7	8.5	1.4
	FAY	2.6	4.7	1.7	3.0	5.4	1.8
	SRB	8.2	15.4	4.4	7.2	13.5	4.2
Fat	DR	4.2	8.1	1.5	3.1	5.8	1.0
	FAY	1.0	2.0	1.2	4.8	9.7	3.7
	SRB	9.3	17.5	5.5	8.7	16.3	4.8
Milk yield	DR	6.8	12.6	1.7	5.0	9.2	1.3
	FAY	4.2	7.8	2.3	4.4	7.9	2.6
	SRB	8.8	16.4	5.1	7.0	13.1	4.2

[1]Gain measures were defined as $\Delta R = \bar{R}_{c,hap} - \bar{R}_{s,SNP}$, the difference between the mean prediction reliabilities; $\Delta Z = \bar{Z}_{c,hap} - \bar{Z}_{s,SNP}$, the difference between Fisher z-transformed data; and

$$\Delta M = \sqrt{\frac{\left(\bar{Z}_{c,hap} - \bar{Z}_{c,SNP}\right)^2}{Var\left(\bar{Z}_{c,hap} - Z_{c,SNP}\right)} + \frac{\left(\bar{Z}_{s,hap} - Z_{s,SNP}\right)^2}{Var\left(\bar{Z}_{s,hap} - Z_{s,SNP}\right)}},$$ the statistical distances.

[2]DR = Danish Red; FAY = Finnish Ayrshire; SRB = Swedish Red.

sible solution, because it allows use of various statistical methods to compare different prediction methods and models. The bootstrap samples made it possible to calculate mean prediction reliabilities and their standard deviations, allowing an approximate assessment of uncertainty.

A notable outcome from the bootstrapping of the test population was the ability to generate confidence ellipses. For both the Bayesian mixture and Bayesian BLUP models, we constructed 95% confidence ellipses for the joint distribution of prediction reliabilities obtained by using SNP or haploblocks, for each training population. We observed a different dispersion of prediction reliabilities within each breed. Except for the prediction of fertility, the dispersion of prediction reliabilities for DR was greater than what was observed for FAY and SRB. One possible explanation of this could be that FAY and SRB are genetically more homogeneous than DR, because the latter could include Holstein genetics (Brøndum et al., 2012). However, a greater dispersion was also observed in prediction of DR using a single-breed training population. Another possible explanation is that the training and test sample sizes of DR were smaller than those of the other 2 breeds (approximately 60 and 40% of the FAY and SRB animals, respectively), increasing the variance of prediction reliabilities over bootstrap samples. Sample size influences variance: the smaller the sample, the smaller the number of individuals sampled in each bootstrap round. This smaller

sample size directly influences reliability estimates, as samples are less representative of the whole population, resulting in more variance of prediction reliabilities over bootstrap samples.

Regarding the training population, the confidence ellipses showed a trend toward an increase in prediction reliabilities when using the combined population instead of the single breed. This increase was statistically significant for predictions of protein, fat, and milk yield in the SRB population, and for predictions of mastitis and fat (only when using Bayesian BLUP) in the FAY population. The increase in prediction reliabilities when using the combined Nordic Red cattle population for training instead of a single breed was as expected, agreeing with what was observed in Brøndum et al. (2011). The dispersion of the results indicated the importance of bootstrapping on the test data set to make reliable statements about the predictions. A single realization of prediction reliability based on the whole test data set may not necessarily reflect the expected result. In some bootstrap samples, prediction reliabilities obtained using the combined training population were smaller than when using a single breed. When looking at the results from the entire set of bootstrap samples, however, although an increase in prediction reliability was not significant for every predicted trait and scenario, a trend toward obtaining larger reliabilities becomes visible when comparing the confidence ellipses.

The benefit from a combined reference data depends on the relationship between populations involved. In the current study, DR, SRB, and FAY have different origins and are 3 different breeds. However, the 3 breeds have exchanged genetic materials in recent decades and have been in a joint breeding program since 2002. Therefore, the gain from the combined reference population is not as large as that obtained from a joint reference data of populations within breed, such as the combined Holstein reference population in EuroGenomics (Lund et al., 2011). However, the gain of the combined reference population in the current study was larger than that from a joint reference data of breeds with weak or no genetic links among them, such as joint reference data combining Holstein and Jersey breeds (Hayes et al., 2009; Erbe et al., 2012; Olson et al., 2012).

The comparison between covariates used to obtain genome-enabled predictions was made by evaluating the shapes of the ellipses. Clear conclusions did not emerge, as most ellipses resembled circles and did not indicate a clear difference between prediction reliabilities obtained with haploblocks or with individual SNP. Therefore, we used ANOVA to support and enrich the analysis. Analysis of variance does not take into account the bivariate structure of the data used for generating the confidence ellipses. Instead, it performs the ANOVA of factors in a regression model that explains prediction reliabilities as a function of covariates (haploblocks or individual SNP), training populations, and their interaction. The *P*-values of the ANOVA indicated that use of haploblocks or SNP was statistically significant mainly when using the Bayesian mixture model. In some cases, the main effect of the covariate type was not significant, but its interaction with the training group was. This provides evidence that use of haploblocks as covariates in across-breed prediction may improve accuracy. The combination of haploblocks with a combined training population resulted in a statistically significant increase of 3.7 to 5.5% in prediction reliabilities, compared with prediction training on a single breed using SNP as covariates. However, haploblocks need to be further explored, to see whether a larger increase in prediction accuracy is feasible. More sophisticated methods for building haploblocks on a multi-breed population and haploblock selection methods could improve the results obtained here.

It is desirable that bootstrap samples be as independent as possible, which is seldom the case because bootstrapping is based on sampling with replacement. All individuals in the test data set had their GEBV obtained from data on the same training individuals, and it was possible for animals to appear in different samples. This is unavoidable when working with bootstrapping. Still, analyses were performed assuming independence of the samples, perhaps increasing the chance of false-positive results.

CONCLUSIONS

We investigated the use of haploblocks as prediction covariates for single- and multi-breed genome-enabled prediction. Results obtained for a combined Nordic Red cattle population indicated that it is possible to increase prediction accuracy over that from methods that use SNP and train on a single breed. Comparisons between prediction reliabilities indicated a clear improvement when using a combined Nordic Red cattle training data set instead of a single breed to train the models. Regarding prediction covariates, we found evidence that haploblocks may improve genomic prediction accuracy over that obtained with individual SNP, but different haplotyping methods and models should be explored to further improve our results. A highlight of this study was the statistical method used to evaluate and compare prediction reliabilities. Bootstrap samples generated from the test data set allowed us to use an innovative multivariate statistical analysis. Concerns due to lack of independence of bootstrap samples were raised, and future studies should address this matter to improve comparison and analysis of prediction reliabilities.

REFERENCES

Boichard, D., F. Guillaume, A. Baur, P. Croiseau, M. N. Rossignol, M. Y. Boscher, T. Druet, L. Genestout, J. J. Colleau, L. Journaux, V. Ducrocq, and S. Fritz. 2012. Genomic selection in French dairy cattle. Anim. Prod. Sci. 52:115–120.

Brøndum, R. F., P. Ma, M. S. Lund, and G. Su. 2012. Short communication: Genotype imputation within and across Nordic cattle breeds. J. Dairy Sci. 95:6795–6800.

Brøndum, R. F., E. Rius-Vilarrasa, I. Strandén, G. Su, B. Guldbrandtsen, 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.

Browning, B. L., and S. R. Browning. 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. Am. J. Hum. Genet. 84:210–223.

Chapman, J. M., J. D. Cooper, J. A. Todd, and D. G. Clayton. 2003. Detecting disease associations due to linkage disequilibrium using haplotype tags: A class of tests and the determinants of statistical power. Hum. Hered. 56:18–31.

Curtis, D. 2007. Comparison of artificial neural network analysis with other multimarker methods for detecting genetic association. BMC Genet. 8:49.

Curtis, D., B. V. North, and P. C. Sham. 2001. Use of an artificial neural network to detect association between a disease and multiple marker genotypes. Ann. Hum. Genet. 65:95–107.

Cuyabano, B. C. D., G. Su, and M. S. Lund. 2014. Genomic prediction of genetic merit using LD-based haplotypes in the Nordic Holstein population. BMC Genomics 15:1171.

Efron, B. 1979. Bootstrap methods: another look at the Jackknife. Ann. Stat. 7:1–26.

Efron, B., and R. Tibshirani. 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy. Stat. Sci. 1:54–77.

Erbe, M., B. J. Hayes, L. K. Matukumalli, S. Goswami, P. J. Bowman, C. M. Reich, B. A. Mason, and M. E. Goddard. 2012. Improving accuracy of genomic predictions within and between dairy cattle breeds with imputed high-density single nucleotide polymorphism panels. J. Dairy Sci. 95:4114–4129.

Fisher, R. A. 1915. Frequency distribution of the values of the correlation coefficient in samples from an indefinitely large population. Biometrika 10:507–521.

Gabriel, S. B., S. F. Schaffner, H. Nguyen, J. M. Moore, J. Roy, B. Blumenstiel, J. Higgins, M. DeFelice, A. Lochner, M. Faggart, S. N. Liu-Cordero, C. Rotimi, A. Adeyemo, R. Cooper, R. Ward, E. S. Lander, M. J. Daly, and D. Altshuler. 2002. The structure of haplotype blocks in the human genome. Science 296:2225–2229.

Gao, H., G. Su, L. Janss, Y. Zhang, and M. S. Lund. 2013. Model comparison on genomic predictions using high-density markers for different groups of bulls in the Nordic Holstein population. J. Dairy Sci. 96:4678–4687.

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

George, E. I., and R. E. McCulloch. 1993. Variable selection via Gibbs sampling. J. Am. Stat. Assoc. 88:881–889.

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

Jairath, L., J. C. M. Dekkers, L. R. Schaeffer, Z. Liu, E. B. Burnside, and B. Kolstad. 1998. Genetic evaluation for herd life in Canada. J. Dairy Sci. 81:550–562.

Janss, L. 2011. BayZ Manual, version 2.04, 2.04 ed. http://www.bayz.biz/.

Johnson, R. A., and D. W. Wichern. 2007. Confidence regions and simultaneous comparisons of component means. Pages 223–229 in Applied Multivariate Statistical Analysis. 6th ed. Pearson Prentice Hall, Upper Saddle River, NJ.

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. Reents, C. Schrooten, F. Seefried, and G. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43.

Mahalanobis, P. C. 1936. On the Generalized Distance in Statistics. National Institute of Science, India.

Meuwissen, T. H. E. 2009. Accuracy of breeding values of "unrelated" individuals predicted by dense SNP genotyping. Genet. Sel. Evol. 41:35.

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.

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.

Patterson, H. D., and R. Thompson. 1971. Recovery of inter-block information when block sizes are unequal. Biometrika 58:545–554.

Schaeffer, L. R. 2001. Multiple trait international bull comparisons. Livest. Prod. Sci. 69:145–153.

Su, G., R. F. Brøndum, P. Ma, B. Guldbrandtsen, G. P. Aamand, and M. S. Lund. 2012. Comparison of genomic predictions using medium-density (∼54,000) and high-density (∼777,000) single nucleotide polymorphism marker panels in Nordic Holstein and Red Dairy Cattle populations. J. Dairy Sci. 95:4657–4665.

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 EAAP, Barcelona, Spain. EAAP, Rome, Italy.

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

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.