Genomic prediction in French Charolais beef cattle using high-density single nucleotide polymorphism markers[1]

M. Gunia,*† R. Saintilan,‡ E. Venot,*† C. Hozé,*†‡ M. N. Fouilloux,§ and F. Phocas*†[2]

*INRA, UMR 1313 Génétique Animale et Biologie Intégrative, 78350 Jouy-en-Josas, France;
†AgroParisTech, UMR 1313 Génétique Animale et Biologie Intégrative, 75231 Paris, France;
‡Union Nationale des Coopératives agricoles d'Elevage et d'Insémination Animale, 149 rue de Bercy,
75595 Paris Cedex 12, France; and §Institut de l'Elevage, 149 rue de Bercy, 75595 Paris Cedex 12, France

ABSTRACT: The objective of the study was to develop a genomic evaluation for French beef cattle breeds and assess accuracy and bias of prediction for different genomic selection strategies. Based on a reference population of 2,682 Charolais bulls and cows, genotyped or imputed to a high-density SNP panel (777K SNP), we tested the influence of different statistical methods, marker densities (50K versus 777K), and training population sizes and structures on the quality of predictions. Four different training sets containing up to 1,979 animals and a unique validation set of 703 young bulls only known on their individual performances were formed. BayesC method had the largest average accuracy compared to genomic BLUP or pedigree-based BLUP. No gain of accuracy was observed when increasing the density of markers from 50K to 777K. For a BayesC model and 777K SNP panels, the accuracy calculated as the correlation between genomic predictions and deregressed EBV (DEBV) divided by the square root of heritability was 0.42 for birth weight, 0.34 for calving ease, 0.45 for weaning weight, 0.52 for muscular development, and 0.27 for skeletal development. Half of the training set constituted animals having only their own performance recorded, whose contribution only represented 5% of the accuracy. Using DEBV as a response brought greater accuracy than using EBV (+5% on average). Considering a residual polygenic component strongly reduced bias for most of the traits. The optimal percentage of polygenic variance varied across traits. Among the methodologies tested to implement genomic selection in the French Charolais beef cattle population, the most accurate and less biased methodology was to analyze DEBV under a BayesC strategy and a residual polygenic component approach. With this approach, a 50K SNP panel performed as well as a 777K panel.

Key words: accuracy, beef cattle, bias, genomic selection, high-density chip

INTRODUCTION

Genomic selection is a way to increase genetic gain by improving the accuracy of the breeding value estimates of young selection candidates that do not necessarily have their own performance record or progeny information. The improved accuracy provides more accurate selection decisions and shorter generation intervals compared to traditional breeding schemes based on bulls progeny testing (Schaeffer, 2006) as in all dairy breeds or in some beef breeds such as the French beef breeds. In dairy breeds, the additional cost of genotyping has been compensated by the progressive abandonment of progeny testing. Similar benefits could also be expected for beef cattle breeds of large population size due to an improved accuracy of all selection candidates and a shorter generation interval for AI bulls. The accuracy of genomic value estimates is the key to successful application of this technology in beef cattle populations. Many factors influence the accuracy of genomic selection, such as size of training population and marker density (Daetwyler et al., 2008; Goddard, 2009; Habier et al., 2013). Many other issues are raised before ap-

Table 1. Heritability (h^2), number of animals in training and validation sets of the full reference population, average reliability of EBV, and average corrected reliability of deregressed EBV (DEBV)[1] for all studied traits

| Trait | h^2 | Training set | | | Validation set | | | Total |
		n	Reliability	Corrected reliability	n	Reliability	Corrected reliability	n
Birth weight	0.36	1,979	0.70	0.60	703	0.53	0.33	2,682
Calving ease	0.09	1,979	0.55	0.40	703	0.37	0.08	2,682
Weaning weight	0.22	1,713	0.60	0.53	670	0.44	0.20	2,383
Muscular development	0.30	1,639	0.63	0.60	676	0.49	0.28	2,315
Skeletal development	0.27	1,639	0.62	0.58	676	0.47	0.25	2,315

[1]Average reliability of the DEBV excluding the parent average contribution.

plying genomic selection, for example, concerning the choice of an accuracy measure (Saatchi et al., 2011), of the response variable and its weight (Garrick et al., 2009), and of a statistical model (de los Campos et al., 2013).

A large training population of about 2,000 animals is now available in French Charolais cattle. Therefore, information from this breed was used to determine the adequate methodology for developing genomic selection in beef cattle populations. The aim of this study was to investigate the technical conditions that are necessary to set up a routine genomic selection in purebred beef cattle in France. The objectives of this research were 1) to estimate the accuracy and bias of genomic value predictions and 2) to document how the density of markers, the size and the structure of the training population, and the prediction methods affect predictive ability.

MATERIALS AND METHODS

Animal Care and Use Committee approval was not obtained for this study because the data were extracted from existing national databases for genetic evaluation purposes.

Genotype Data

A total of 2,751 registered French Charolais bulls and cows were genotyped either with the Bovine SNP50 BeadChip (50K) for 2,079 animals or with the BovineHD BeadChip (777K) for the 672 main genetic contributors to the Charolais breed. The cryopreserved semen or ear samples used as material for DNA extraction were procured by various AI organizations, the Charolais breeder association, and INRA. Animals with genotypes inconsistent with pedigree information were discarded.

A quality control of SNP genotypes based on call rate (90%) and Hardy Weinberg equilibrium test (P-value $< 10^{-4}$) was performed in the same manner on the 50K and 777K genotypes. The SNP were mapped to the UMD 3.1 build of the bovine genome sequence assembled by the Center of Bioinformatics and Computational Biology at the University of Maryland (College Park, MD).

After quality control, 708,771 SNP of the 777K SNP chip were retained on 664 animals and 45,187 SNP of the 50K SNP chip were retained on 2,078 animals. Imputation of the 50K genotype data to 777K genotypes was performed using BEAGLE software (Beagle Software, Minneapolis, MN) for these 2,078 animals (Browning and Browning, 2009). A detailed description of genotype editing and imputation procedure is given by Hozé et al. (2013).

In total, 2,742 true or imputed genotypes were available for the study. Animals without their own performance or progeny records were excluded from the study leaving 2,682 animals (94% males) for analysis.

Phenotype Data

Five field traits on which national genetic evaluations exist were considered in this study: birth weight, calving ease, weaning weight, muscular development, and skeletal development.

Information on the number of records and reliability of estimated breeding values for each trait are presented in Table 1 for the full reference population. We tested different scenarios (described below) with different training sets. In the reference scenario, 1,979 animals born between 1965 and 2011 were used in the training set to estimate SNP effects. Across all scenarios, the same sample consisting of 703 animals born in 2012 was used for validation. The animals in the validation set had records only on their own performance and not on progeny. The birth year distribution of the reference population according to the availability of progeny records is presented in Fig. 1.

Two kinds of response variables were used: EBV and deregressed EBV (**DEBV**) from traditional BLUP genetic evaluation. They were considered in weighted analyses to account for heterogeneous variances of the response variable due to a variable amount of progeny records among genotyped animals: the weight of EBV in the analysis was their accuracy (square root of their reliability) and the weight of DEBV were derived according to the method proposed by Garrick et al. (2009). The deregression
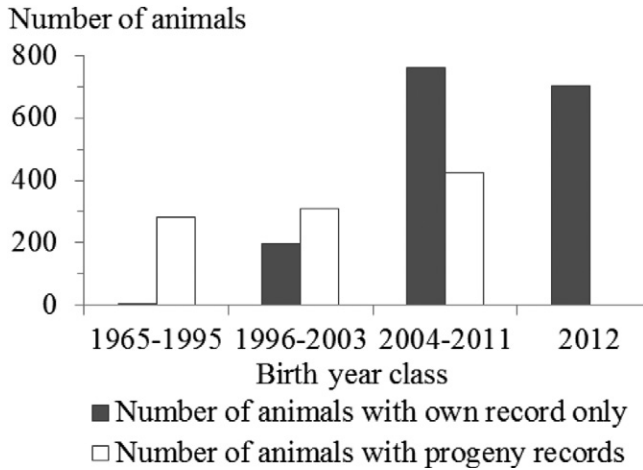
Number of animals

Figure 1. Number of animals with own or progeny records for birth weight in the training set (1965 to 2011) and validation set (2012).

method removed parent average both from the response variable and the weight (Garrick et al., 2009).

Statistical Models

Description. Genomic prediction equations were derived using 2 different methods: genomic BLUP (**GBLUP**; VanRaden, 2008) and BayesCπ (Habier et al., 2011). In these methods, the priors of all SNP effects are assumed to share the same variance. With regard to GBLUP, the prior used is a normal distribution. Concerning BayesCπ approach, the prior is a mixture distribution. The effects of SNP are fitted with a probability π, where π is the fraction of SNP having an effect. Here, convergence for the estimation of parameter π was not obtained. Therefore, we assumed the parameter known and moved to a BayesC strategy (Kizilkaya et al., 2010). A BayesC approach is known to provide good results in genomic selection (Croiseau et al., 2012). The π parameter was fixed at a value of 0.001 (i.e., 708 markers with a nonzero effect) when using the 777K chip and 0.0157 (i.e., 711 markers with nonzero effect) when using the 50K chip. A traditional BLUP, hereafter named pedigree-based BLUP, was also considered to estimate nongenomic breeding values based on pedigree and provide a basis for comparison.

Analyses were performed using the GS3 software (Legarra et al., 2013). Initial variances were estimated with GS3 with the VCE option. For BayesC, a burn-in period of 20,000 iterations was run before saving results every 10 iterations out of 50,000. A greater number of iterations was also tested (300,000). Results in terms of variances and accuracies were found to be similar, so only 50,000 iterations were run for all the analyses.

General Model. For each trait, the following model was fit to the response variable y (EBV or DEBV) for the training populations:

$$y = \mathbf{1}\mu + \mathbf{Ma} + \mathbf{e},$$

in which **1** is a vector of 1, μ is the overall mean, and **M** is an incidence matrix for marker genotypes. The genotypes were coded as –1, 0, or 1 depending on the number of copies of a given marker allele carried by the animal, **a** is a vector of marker effects, and **e** is a vector of residual effects.

Once the marker effects were estimated with either GBLUP or BayesC methods, the predicted genomic value (genomic EBV [**GEBV**]) of an individual was

$$\mathrm{GEBV}_i = \sum_{j=1}^{J} M_{ij} \widehat{a}_j,$$

in which GEBV_i is the GEBV for animal i in the validation population, M_{ij} is the marker genotype of animal i at marker j, J the total number of markers, and $\widehat{a}_j$ the effect of marker j.

Inclusion of a Polygenic Component. We also considered a variation of the general model by including a residual polygenic component in the analysis. The total genetic variance was partitioned into 2 components: the additive genetic variance explained by the markers and the residual polygenic variance. We fixed the residual polygenic variance at different fractions of the total genetic variance of the trait. According to Garrick et al. (2009), there are 2 ways to account for the residual polygenic component in the analysis: either by accounting for a polygenic fraction of the genetic variance in the weights (w_i) or by explicitly including a polygenic component in the model.

To consider the polygenic fraction in the weights of the DEBV, the formula derived by Garrick et al. (2009) was used:

$$w_i = \left(1 - \mathrm{h}^2\right) / \left\{ c + \left[\left(1 - r_i^2\right) / r_i^2 \right] \right\} \mathrm{h}^2,$$

in which r_i^2 is the reliability of DEBV, h^2 is the heritability of the trait, and c is the polygenic fraction of the genetic variance.

When the residual polygenic component was explicitly included in the model, the equation became

$$y = \mathbf{1}\mu + \mathbf{Ma} + \mathbf{Zu} + \mathbf{e},$$

in which **u** is a vector of polygenic effects and **Z** is an incidence matrix for the polygenic effects.

For the latter model, the predicted genomic value for an individual was defined as the sum of the predicted effects of the SNP over all the markers and the polygenic breeding value:

$$\widehat{\mathrm{GEBV}}_l = \sum_{j=1}^{J} M_{ij} \widehat{a}_j + \widehat{u}_i.$$

Table 2. Presentation of the main features of the tested scenarios

Main features	Name of the scenarios				
	"Reference"	"Model and marker density"	"3 reduced training populations: the 50% youngest, the 50% oldest, progeny tested animals only"	"EBV as response variable"	"Residual polygenic component"
Model and marker density[1]	BayesC 777K	BLUP BayesC GBLUP 50K, 777K	BayesC 777K	BayesC 777K	BayesC 777K
No. of animals in the training set	1,979	1,979	The 50% youngest: 990 The 50% oldest: 989 Progeny tested: 1,016	1,979	1,979
Response variable	DEBV[2]	DEBV	DEBV	EBV	DEBV
Residual polygenic component	No	No	No	No	Yes: in the model or in the weights

[1]50K = 50,000 SNP panel; 777K = 777,000 SNP panel; GBLUP = genomic BLUP.

[2]DEBV = deregressed EBV.

Validation Criteria

The accuracy of GEBV could be defined as the correlation between true genetic values and GEBV. The true genetic values of genotyped animals are not available in real data sets and consequently an approximation must be used to assess the accuracy of GEBV. A good approximation of the true breeding value is DEBV when the animals in the validation set have records over many progenies. However, in beef cattle validation populations, it may not be the general case. In our situation, this is even far from reality because the validation population is only known for own performance and DEBV is then similar to own phenotype. Therefore, another approximation was also considered, which is the traditional BLUP EBV.

To assess the predictive ability of genomic equations, the accuracy of genomic prediction was estimated for the validation population in 3 different ways:

- as the Pearson's correlation between DEBV and GEBV,

- as this previous correlation divided by the square root (h) of the h^2 because the expectation of the previous correlation is h, and

- as the Pearson's correlation between EBV and GEBV assuming that the true breeding value is approximated by EBV.

To evaluate the bias of genomic predictions, the regression coefficients of the response variables (DEBV or EBV) on GEBV were derived and compared to their expected value of 1.

Scenarios

We tested different scenarios listed in Table 2.

1) The reference scenario was used as a basis for comparison. It features a BayesC model with the 777K SNP panels, with a full data set and DEBV as response variable.

2) Variations of the reference scenario were tested by changing the model and marker density (BayesC versus GBLUP with 777K or 50K SNP) and comparing with pedigree-based BLUP.

3) Three scenarios with half the training population were tested to assess the impact of birth year and information available for animals in the training set on the accuracy of genomic selection. The original training population was divided into 2 with the 50% youngest animals (born between 2004 and 2011), and the 50% oldest animals (born between 1965 and 2003). Each half dataset was used to estimate the SNP effects. A third training set was created by excluding animals having only own performance records. Only progeny tested animals were kept; they represented 51% of the complete training population. The different training populations are described in Table 3. Their average relationship coefficients within populations and between the training and the common validation populations are presented in Table 4. Out of the 703 young animals in the validation set, 645 were sired by 77 bulls of the training set.

4) The scenario with EBV as a response variable features EBV instead of DEBV as phenotype to estimate the SNP effects.

5) In the last scenarios, a residual polygenic component, representing different fractions of the total genetic variance was considered either in the statistical model or through the weights.

Table 3. Number of animals and the average corrected reliability of their deregressed EBV (DEBV)[1] for 3 reduced training populations: "the 50% youngest," "50% oldest,"[2] and "progeny tested"[3]

Trait	The 50% youngest		The 50% oldest		Progeny tested animals		
	n	Reliability	n	Reliability	n	Reliability	n[4] of progeny
Birth weight	990	0.50	989	0.70	1,016	0.86	1,314
Calving ease	990	0.27	989	0.53	1,016	0.70	1,313
Weaning weight	864	0.38	949	0.67	926	0.81	862
Muscular development	820	0.46	819	0.73	929	0.84	864
Skeletal development	820	0.43	819	0.72	929	0.83	864

[1]Average reliability of the DEBV excluding the parent average contribution.

[2]The 50% youngest animals and the 50% oldest animals of the original training population.

[3]A training population excluding animals having only own performance records.

[4]nb = Average number of progeny per animal.

RESULTS

Assessing the Accuracy of Genomic Predictions

In our first analysis based on DEBV response variable, we compared different criteria to assess the accuracy of genomic prediction for the validation population. They are presented in Table 5 for a BayesC model with 777K or 50K genotypes. Whatever the criterion considered, accuracy of genomic prediction under BayesC was very similar between 777K genotypes or 50K genotypes. The greatest values of accuracy were obtained with the correlations between traditional BLUP EBV and GEBV, whereas estimates of accuracy based on the correlations between DEBV and GEBV had the lowest values.

The best choice of criteria to assess the accuracy would depend on what is the best approximation of the true breeding value and on the degree of independence between training and validation populations. In the next sections, we will only present the accuracy derived as the correlation between DEBV and GEBV divided by the square root of heritability to provide similar criterion of accuracy as in some other publications in beef cattle (Saatchi et al., 2011; Elzo et al., 2012; Bolormaa et al., 2013).

Different Models and Marker Densities

We compared the accuracies and regression coefficients of pedigree-based BLUP, GBLUP, and BayesC in Table 6. The scenarios with BayesC had a greater accuracy than with GBLUP (of 0.05 to 0.08 depending on the marker density) and BLUP (of 0.06). When shifting from the 50K to the 777K SNP chip, no gain in accuracy was observed with BayesC and a loss of accuracy and an increase of bias was observed with GBLUP. Pedigree-based BLUP showed less biased estimates and greater accuracies for birth weight and calving ease but was outperformed by genomic methods for the other traits.

Different Training Sets and Response Variables

Table 7 presents the accuracies and regression coefficients of genomic predictions for 3 different reduced training population sizes (the oldest 50%, the youngest 50%, and progeny tested animals).

Halving the training set and taking only the oldest 50% of the animals brought a slightly greater accuracy than taking only the youngest 50% (Table 7). The slight increase in accuracy depending on the birth periods of training animals is due to greater reliabilities of DEBV of the oldest bulls compared to the youngest (Table 3), which is partially compensated by a greater average relationship between the youngest training population and the validation population (Table 4).

The regression coefficients of DEBV on GEBV for muscular and skeletal development showed less variability across scenarios than other traits. The regression coefficient of muscular development was close to 1 and therefore showed very little bias. The GEBV of birth weight, calving ease, and skeletal development were inflated in all scenarios (the regression coefficients of DEBV on GEBV were lower than 1). The GEBV for weaning weight were underestimated for all scenarios except for the one with the youngest 50% training population where the GEBV was inflated.

The difference in accuracy between the full training set and the training set containing only animals with

Table 4. Average relationship coefficients[1] and their SD within and between groups for the 4 training populations and the common validation population

Population	Within group	Between training and validation set
Full training population	0.020 ± 0.032	0.026 ± 0.028
50% youngest animals	0.026 ± 0.034	0.028 ± 0.030
50% oldest animals	0.018 ± 0.036	0.022 ± 0.026
Progeny tested animals	0.026 ± 0.036	0.028 ± 0.030
Validation population	0.040 ± 0.048	

[1]Average relationship coefficients derived from the pedigree.

Table 5. Three different criteria to assess accuracy of genomic prediction under a BayesC model applied to the complete training population genotyped either for 777K or 50K[1]

Trait	Pearson's correlation between DEBV and GEBV[2]		Pearson's correlation between DEBV and GEBV divided by the square root of heritability		Pearson's correlation between EBV and GEBV	
	777K	50K	777K	50K	777K	50K
Birth weight	0.25	0.23	0.42	0.38	0.53	0.50
Calving ease	0.10	0.10	0.34	0.34	0.64	0.63
Weaning weight	0.21	0.21	0.45	0.45	0.65	0.63
Muscular development	0.29	0.30	0.52	0.54	0.59	0.59
Skeletal development	0.14	0.14	0.27	0.27	0.52	0.52
Mean	0.20	0.20	0.40	0.40	0.59	0.57

[1]777K = 777,000 SNP panel; 50K = 50,000 SNP panel.

[2]DEBV = deregressed EBV; GEBV = genomic EBV.

progeny records was 0.02 on average, with the largest increase occurring for birth weight (+0.05; Table 7) in the full data set. This means that doubling the training population by adding animals having only their own performance recorded brought little additional information.

The accuracies were similar when using DEBV or EBV as response variables to fit the genomic prediction equations, with a slightly greater accuracy on average for DEBV (+0.02; Table 7) due to a clearer advantage of DEBV as response variable for muscular and skeletal development scores only. Bias was slightly reduced when using EBV.

Table 6. Accuracy[1] and regression coefficient[2] of predicted EBV[3] under BayesC, genomic BLUP (GBLUP),[4] and BLUP models

Trait	BayesC		GBLUP		BLUP
	777K	50K	777K	50K	–
Accuracy					
Birth weight	0.42	0.38	0.33	0.40	0.45
Calving ease	0.34	0.34	0.25	0.21	0.41
Weaning weight	0.45	0.45	0.42	0.45	0.40
Muscular development	0.52	0.54	0.34	0.38	0.23
Skeletal development	0.27	0.27	0.26	0.31	0.22
Mean	0.40	0.40	0.32	0.35	0.34
Regression coefficient					
Birth weight	0.71	0.64	0.55	0.74	1.15
Calving ease	0.53	0.54	0.40	0.36	0.91
Weaning weight	1.13	1.09	0.97	1.12	1.19
Muscular development	0.95	1.00	0.69	0.81	0.61
Skeletal development	0.64	0.64	0.63	0.82	0.69
Mean	0.79	0.78	0.65	0.77	0.91
Average deviation[5] from 1	0.26	0.25	0.35	0.28	0.23

[1]Accuracy is measured by Pearson's correlation between observed deregressed EBV (DEBV) and predicted EBV in the validation population divided by the square root of heritability.

[2]Regression coefficient of observed EBV on predicted DEBV. A coefficient of 1 is expected.

[3]Predicted EBV are GEBV for BayesC and GBLUP methods.

[4]BayesC and GBLUP were run with 777K (777,000) and 50K (50,000) SNP panels.

[5]Average absolute deviation of the regression coefficient from 1.

Table 7. Accuracy[1] and regression coefficient[2] of predicted genomic EBV (GEBV) under BayesC model with 777K SNP panel[3] for the reference scenario, the scenarios with 3 different reduced training population sizes, and the scenario with EBV as a response variable

Trait	Reference scenario	Reduced training population sizes			EBV as response variable
		The 50% youngest	The 50% oldest	Progeny tested animals	
Accuracy					
Birth weight	0.42	0.37	0.35	0.37	0.42
Calving ease	0.34	0.15	0.32	0.32	0.33
Weaning weight	0.45	0.37	0.35	0.44	0.45
Muscular development	0.52	0.37	0.41	0.51	0.47
Skeletal development	0.27	0.17	0.22	0.27	0.22
Mean	0.40	0.28	0.33	0.38	0.38
Regression coefficient					
Birth weight	0.71	0.90	0.71	0.64	0.82
Calving ease	0.53	0.44	0.55	0.52	0.60
Weaning weight	1.13	1.51	0.88	1.11	1.04
Muscular development	0.95	0.93	0.92	1.01	1.01
Skeletal development	0.64	0.63	0.61	0.66	0.52
Mean	0.79	0.88	0.74	0.79	0.80
Average deviation[4] from 1	0.26	0.32	0.26	0.26	0.22

[1]Accuracy is measured by the Pearson's correlation between deregressed EBV (DEBV) and GEBV divided by the square root of heritability.

[2]Regression coefficient of DEBV on GEBV.

[3]777K = 777,000 SNP panel.

[4]Average absolute deviation of the regression coefficient from 1.

Table 8. Accuracy[1] and regression coefficient[2] of predicted genomic EBV (GEBV) under BayesC model with 777K SNP panel[3] for the reference scenario (polygenic fraction of 0%) and the scenario with a residual polygenic component in the model (or in brackets average results for a residual polygenic component integrated through the weights of the deregressed EBV [DEBV] response variable)

Polygenic fraction of the genetic variance	0%	10%	20%	30%	40%	50%
Accuracy						
Birth weight	0.42	0.43	0.44	0.46	0.48	0.47
Calving ease	0.34	0.34	0.37	0.38	0.40	0.41
Weaning weight	0.45	0.45	0.45	0.45	0.45	0.45
Muscular development	0.52	0.52	0.52	0.51	0.50	0.49
Skeletal development	0.27	0.27	0.27	0.27	0.27	0.27
Mean	0.40	0.40	0.41	0.42	0.42	0.42
(in the weights)	(0.40)	(0.40)	(0.40)	(0.40)	(0.39)	(0.38)
Regression coefficient						
Birth weight	0.71	0.77	0.83	0.91	0.99	1.02
Calving ease	0.53	0.57	0.63	0.69	0.76	0.80
Weaning weight	1.13	1.15	1.19	1.22	1.23	1.25
Muscular development	0.95	0.99	1.01	1.02	1.03	1.02
Skeletal development	0.64	0.68	0.70	0.73	0.76	0.77
Mean	0.79	0.83	0.87	0.92	0.95	0.97
(in the weights)	(0.79)	(0.95)	(1.01)	(1.21)	(1.32)	(1.41)
Average deviation[4] from 1	0.26	0.23	0.21	0.18	0.15	0.14

[1]Accuracy is measured by the Pearson's correlation between DEBV and GEBV divided by the square root of heritability.

[2]Regression coefficient of DEBV on GEBV.

[3]777K = 777,000 SNP panel.

[4]Average absolute deviation of the regression coefficient from 1.

Polygenic Component

Accuracies and regression coefficients were presented in Table 8 when a polygenic component was included in the model or in the weights of the response variable. Accuracies were more robust than regression coefficients to changes of the residual polygenic fraction. The accuracy increased slightly for birth weight and calving ease when a polygenic component was included. This increase was slightly greater when the polygenic component was included in the model than in the weights. For the other traits, accuracy remained constant or even decreased slightly when considering a polygenic component either in the model or in the weights.

Bias was reduced when considering a residual polygenic component for birth weight, calving ease, and skeletal development. It was also reduced for muscular development but only when the polygenic component was included in the model. Reduction of the bias concerned traits for which the regression coefficient of DEBV on GEBV is lower than 1. In the case where the pure genomic model led to a likely underestimation of the GEBV (i.e., regression coefficient of DEBV on GEBV > 1 of weaning weight), including a polygenic component in the analysis worsened the underestimation, especially when considering high (>20%) residual polygenic component in the weights of the response variable.

The optimal percentage of polygenic variance seems to vary across traits and methods, from 0% for weaning weight to 50% for calving ease (with both methods of accounting for a residual polygenic component). These observations are in accordance with the accuracy obtained for a 100% polygenic model (BLUP in Table 6), which gave a greater accuracy for calving ease and lower accuracy for weaning weight compared to genomic methods.

The other traits are on an intermediary scale. For birth weight, the optimal percentage of polygenic variance in the model was 40% when the polygenic fraction was included in the model and 20% when it was included in the weights of the response variable. For muscular development, it was 10 or 20% in the model and 0% in the weights. For skeletal development, it was 50% in the model and 30% in the weights.

DISCUSSION

Statistical Methods

Our results show that the method with BayesC and a residual polygenic component is the best approach tested for genomic selection in the French Charolais seedstock population. The density of markers (between 50K and 777K) is of little importance with BayesC. On the opposite, GBLUP was less accurate and more biased with 777K marker density than with 50K, probably because the meth-

od considers that all the markers have an effect, resulting in a greater prediction error for a higher number of SNP effects to estimate from the same small training population. BayesC performed better than GBLUP, probably because only a small proportion of the markers were considered having an effect in BayesC and because this methodology takes linkage disequilibrium (**LD**) better into account. These results are consistent with previous studies (Erbe et al., 2012; Pryce et al., 2012). Erbe et al. (2012) observed a decrease in accuracy with GBLUP when the 777K panel was used rather than the 50K panel within breed. As stated by Erbe et al. (2012), methods that remove SNP from the model or set their effects to 0 are necessary to take advantage of the increased marker density.

The pattern of accuracy across the different methods and polygenic fractions suggests different genetic architectures of traits. Best linear unbiased prediction or genomic models with a high polygenic component gave better results for calving ease, suggesting that this trait is determined by many loci with small effects. On the contrary, weaning weight may be determined by fewer loci with bigger effects as suggested by the better results obtained without a polygenic fraction.

Assessing the Accuracy of Genomic Predictions

In our study, estimates of accuracy based on the correlation between DEBV and GEBV were low because DEBV correspond only to the own phenotype of animals in the validation population. Consequently, even if the predictive ability of the genomic equations to estimate breeding values were close to 1, the expected value of this correlation would only be the square root of the heritability. Estimates of accuracy based on the correlation between BLUP EBV and GEBV were the greatest. Two different points may explain the latter result: the first one is that the BLUP EBV is the best predictor of the true breeding value based on pedigree and observed phenotypes; the second one is because the BLUP EBV are based on pedigree and observed phenotypes and SNP capture well pedigree relationships (Habier et al., 2007). In our case, most of the animals in the validation population have their sires or maternal grandsires in the training population; using their EBV in the validation criterion is therefore not recommended because it will capture the pedigree information. In case the validation population is not closely pedigree linked to the training population, using EBV as the best predictor of true breeding values to assess accuracy of genomic predictions should give a good approximation of the true accuracy.

No agreement was found in the literature on a common way to approximate accuracy of genomic predictions from real data. Since true breeding values are never known, different criteria were used as a surrogate for the true accuracy defined as the correlation between the GEBV and the true breeding values. The best choice of criterion to assess the accuracy would depend on the phenotypic information available for the validation population and on their relationship coefficients with the training population. Simple correlations between phenotypic response variables and GEBV were mostly used in dairy cattle (Hayes et al., 2009; Brondum et al., 2012). In dairy cattle, this is a very good approximation of true accuracy, because bulls included in validation populations have records on tens or hundreds of daughters, and consequently the response variable is a good prediction of the true breeding value. When the response variable for the validation population is mainly based on own phenotype with no or little progeny information, it has been proposed to approximate the accuracy as the correlation between response variable and GEBV divided by the square root of heritability (Daetwyler et al., 2012; Bolormaa et al., 2013). Some authors also reported accuracy measured by the coefficient of determination of the regression of the response variable on the GEBV (Karoui et al., 2012; Olson et al., 2012).

Reliability defined as the squared correlation between the phenotype measure and GEBV was also used (Su et al., 2010). In some studies this squared correlation was divided by the weighted mean reliability of the response variable (Liu et al., 2011; Lund et al., 2011; Thomasen et al., 2012) to account for the fact that the response variable is only an approximation of the true breeding value.

Training and Validation Population Definition

In cattle breeding schemes, GEBV should be predicted for young selection candidates without phenotypes but sired by genotyped and phenotyped bulls. To be as close as possible of a realistic genomic selection program, we chose a validation population constituted of young animals that had their sires in the training population. Such strategy is the one that is promoted and commonly used in dairy cattle validation studies (Lund et al., 2011). However, it is well known that close relationships between animals in the training and validation sets increases the accuracy of genomic predictions compared to the ones derived for an independent validation population (Habier et al., 2011). A loss of accuracy of 16% on average is expected in the French Charolais beef cattle when comparing strategies with a validation population without sires in training populations of same sizes (T. Tribout, INRA, Jouy-en-Josas, France, personal communication).

Own Performance Records

The relationships between animals in the training and validation sets were high. Therefore, adding to the training population animals that have lower relationship with the validation animals could not greatly improve the accuracies.

The added animals with own performance only were typically females or young bulls. The accuracy of the response variable is expected to increase when the performance of their progeny is recorded. However, the accuracy of female response variable will never reach the level of bulls. The low accuracy of female information is the reason why most of the reference populations in cattle were built based on male information only (Karoui et al., 2012; Su et al., 2012).

However, genotyping females is especially relevant for small cattle breeds that only have a limited number of progeny tested bulls (Jimenez-Montero et al., 2012). It is also a growing concern in dairy breeds of large population size, for recording new traits (Buch et al., 2012), to gain accuracy (Tsuruta et al., 2013), or because not enough progeny tested bulls are available (Ding et al., 2013). Tsuruta et al. (2013) observed a gain in reliability of 2 to 3% in U.S. Holstein cattle when using female genotypes in addition to male genotypes.

Population sizes of beef cattle breed stocks are much smaller than that of the Holstein breed, and AI is much less used, resulting in a lower availability of reliable sires for training population (Garrick, 2011). Therefore, other strategies are needed in beef cattle to obtain a large reference population such as genotyping females, pooling reference populations across countries, or eventually pooling reference populations across breeds.

Response Variables

When comparing EBV and DEBV as a response variable, we found a slight advantage for DEBV in terms of accuracy. In a simulation approach, Guo et al. (2010) compared daughter yield deviation (**DYD**) and EBV as response variables. They showed that EBV approach performed as well as DYD or better than the DYD approach in terms of reliability of predictions. It was especially true for traits with low heritability or training populations where bulls had a low number of daughters. Guo et al. (2010) observed that DEBV were theoretically superior to EBV with regard to double counting and double regression, but this advantage was counteracted by less information and more random errors in DEBV. Estimated breeding values are predicted from data of all available relatives and therefore contain relatively little random error and high reliability.

Gredler et al. (2010) found also a slightly greater accuracy by using EBV than DEBV or DYD in genomic predictions for Fleckvieh cattle. On the opposite, Ostersen et al. (2011) found greater accuracy using DEBV than EBV in pig.

Garrick et al. (2009) proposed a method to deregress EBV and remove parent average effects to address the issues raised by EBV. The authors emphasize the importance of using DEBV instead of EBV to eliminate the shrinkage feature of the BLUP EBV and to avoid double counting of relatives' information in the genomic predictions. They also mentioned that prediction errors of EBV are negatively correlated with the true breeding values.

On the basis of our results alongside those of the literature, the choice of the response variable seems to have little impact on accuracy of genomic prediction.

Polygenic Component

In our study, including a polygenic component was favorable for all traits except weaning weight. The optimal fraction of residual polygenic variance varied across traits and methods (polygenic fraction in the model or in the weights).

Liu et al. (2011) tested a residual polygenic effect included in the model, whose variance was representing different percentages of the total genetic variance. They also observed that according to the regression coefficients, the optimal percentage of residual polygenic variance seems to vary across traits.

The residual polygenic component can either be included in the model (Calus and Veerkamp, 2007) or in the weights of the response variable (Garrick et al., 2009). The proportion of additive genetic variance not explained by the markers is not known before the training analyses. If the polygenic component is explicitly included in the model, the polygenic variance can be estimated during the training analysis. If the polygenic component is included in the weights, the value of c can be estimated from a first validation analysis and the training analysis could then be repeated using the estimated value of c (Garrick et al., 2009). These authors alternatively suggested assessing the sensitivity of results to the c value by using a range of values.

Including a polygenic term is done in the French marker-assisted selection program (Guillaume et al., 2008; Boichard et al., 2012) but is not always done in genomic prediction analysis (Garrick et al., 2009) because genomic models generally assume that SNP explain all the genetic variation (Meuwissen et al., 2001). Some authors found that including a polygenic term reduces bias of GEBV (Liu et al., 2011) and bias of SNP or haplotype variances (Calus and Veerkamp, 2007; Rius-Vilarrasa et al., 2012). Including a polygenic term increases the persistency of accuracy and the stability of regression coefficient over generations (Solberg et al., 2009), and the models are less sensitive to the prior assumption about marker effects (Rius-Vilarrasa et al., 2012). Some authors observed a slight reduction of the accuracy (Liu et al., 2011; Rius-Vilarrasa et al., 2012). We also observed a reduction of bias for most of the traits and a decrease in accuracy for some traits.

Inclusion of a residual polygenic component seems to be more important for low heritable traits and low marker density (Calus and Veerkamp, 2007; Duchemin et al., 2012); these authors observed that the polygenic variance explained a greater proportion of the estimated genetic variance and bring a higher accuracy.

Different reasons justify the advantages of using a polygenic component. The inclusion of a polygenic component allows selecting QTL with rare alleles (Goddard, 2009) and capturing the variance of QTL with small effects (Calus and Veerkamp, 2007), thus reducing bias. Estimates of polygenic effects are based on BLUP theory and therefore show little bias (Solberg et al., 2009), which also contribute to bias reduction. Moreover, the persistence of accuracy over time is greater when including a polygenic component because the remaining marker association reflects LD more truly (Solberg et al., 2009). Therefore, fitting a polygenic component, either in the model or in the weight of the response variable, could be advised for some of the traits in Charolais beef cattle. However, the benefits of fitting a polygenic term can be achieved if own records or records on relatives of the selection candidates are available, which is not always the case in beef cattle populations.

Comparison with Genomic Selection Implemented in other Beef Breeds

It is difficult to compare the accuracy of genomic predictions across studies because of the different genetic structures and training sizes of populations, models, and validation methods used.

The accuracy of GEBV in beef cattle populations are expected to be less than those typically found in Holstein breed (Erbe et al., 2012; Colombani et al., 2013) even for the same reference population sizes. This can be explained by larger effective population sizes and lower accuracies of bull EBV due to a low use of artificial insemination in beef breeds compared to dairy breeds.

The accuracy measured by the simple correlation between DEBV and GEBV in Charolais is lower than the one observed for American Hereford cattle (Saatchi et al., 2013), although the training Hereford population was only composed of 772 bulls. In Charolais, we reported correlations of 0.25 for birth weight, 0.10 for calving ease, and 0.21 for weaning weight. In American Hereford, they reported correlations of 0.37 for birth weight, 0.25 for calving ease, and 0.51 for weaning weight with a BayesC model and validation on the youngest animals. The lower correlations in Charolais compared to American Hereford is probably partly due to the greater effective population size in French Charolais of about 500 (Bouquet et al., 2011) than the effective size of American Hereford population of 85 (Cleveland et al., 2005). However, the main explanation could be the difference in the type of records used for the validation population, but information was not fully detailed in Saatchi et al. (2013) to confirm this hypothesis. In our Charolais study, the validation population concerned only animals without offspring's records, whose DEBV had a low reliability. The simple correlation between DEBV and GEBV was therefore a very strong underestimation of the correlation between true breeding value and GEBV. The only study that allows a fair comparison with our results is from Saatchi et al. (2011) on American Angus because their American Angus population has both an effective size and a training population size close to our French Charolais population. The large training population of 2,500 Angus bulls with average reliability of DEBV of 0.8 and 0.7 for birth and weaning weights, respectively, has to be compared to the 2,000 Charolais bulls with average reliability of 0.6 and 0.5 for birth and weaning weights, respectively. Saatchi et al. (2011) assessed the accuracy by the correlation between DEBV and GEBV divided by the square root of heritability. Transforming their results to simple correlation between DEBV and GEBV, they reported accuracies of 0.33 for birth weight and 0.25 for weaning weight under a BayesC and validation on youngest animals, which are greater than for Charolais, with 0.25 and 0.21 for birth and weaning weights, respectively.

Conclusion

Among the approaches tested, the methodology that appears to be the most accurate and less biased to implement genomic selection in a purebred beef cattle population such as the French Charolais population is to use DEBV as response variable under a BayesC genomic selection strategy. Adding a residual polygenic component in the analysis reduces the bias of GEBV for most of the traits. Using a 777K SNP panel instead of a 50K panel does not give a clear advantage for increasing the accuracy of genomic predictions within Charolais breed. In addition, genotyping more animals to increase the reference population should be carefully considered as animals with only own performance bring little gain in prediction accuracy.

LITERATURE CITED

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.

Bolormaa, S., J. E. Pryce, K. Kemper, K. Savin, B. J. Hayes, W. Barendse, Y. Zhang, C. M. Reich, B. A. Mason, R. J. Bunch, B. E. Harrison, A. Reverter, R. M. Herd, B. Tier, H. U. Graser, and M. E. Goddard. 2013. Accuracy of prediction of genomic breeding values for residual feed intake and carcass and meat quality traits in *Bos taurus*, *Bos indicus*, and composite beef cattle. J. Anim. Sci. 91:3088–3104.

Bouquet, A., E. Venot, D. Laloe, F. Forabosco, A. Fogh, T. Pabiou, K. Moore, J. A. Eriksson, G. Renand, and F. Phocas. 2011. Genetic structure of the European Charolais and Limousin cattle meta-populations using pedigree analyses. J. Anim. Sci. 89:1719–1730.

Brondum, R. F., G. Su, M. S. Lund, P. J. Bowman, M. E. Goddard, B. J. Hayes, and G. S. Su. 2012. Genome position specific priors for genomic prediction. BMC Genomics 13:543.

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.

Buch, L. H., M. Kargo, P. Berg, J. Lassen, and A. C. Sorensen. 2012. The value of cows in reference populations for genomic selection of new functional traits. Animal 6:880–886.

Calus, M. P. L., and R. F. Veerkamp. 2007. Accuracy of breeding values when using and ignoring the polygenic effect in genomic breeding value estimation with a marker density of one SNP per cM. J. Anim. Breed. Genet. 124:362–368.

Cleveland, M. A., H. D. Blackburn, R. M. Enns, and D. J. Garrick. 2005. Changes in inbreeding of U.S. Herefords during the twentieth century. J. Anim. Sci. 83:992–1001.

Colombani, C., A. Legarra, S. Fritz, F. Guillaume, P. Croiseau, V. Ducrocq, and C. Robert-Granie. 2013. Application of Bayesian least absolute shrinkage and selection operator (LASSO) and BayesC pi methods for genomic selection in French Holstein and Montbeliarde breeds. J. Dairy Sci. 96:575–591.

Croiseau, P., F. Guillaume, and S. Fritz. 2012. Comparison of genomic selection approaches in Brown Swiss within intergenomics. Interbull Bull. 46:127–132.

Daetwyler, H. D., A. A. Swan, J. H. van der Werf, and B. J. Hayes. 2012. Accuracy of pedigree and genomic predictions of carcass and novel meat quality traits in multi-breed sheep data assessed by cross-validation. Genet. Sel. Evol. 44:33.

Daetwyler, H. D., B. Villanueva, and J. A. Woolliams. 2008. Accuracy of predicting the genetic risk of disease using a genome-wide approach. PLoS ONE 3:e3395.

de los Campos, G., J. M. Hickey, R. Pong-Wong, H. D. Daetwyler, and M. P. L. Calus. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193:327–345.

Ding, X., Z. Zhang, X. Li, S. Wang, X. Wu, D. Sun, Y. Yu, J. Liu, Y. Wang, Y. Zhang, S. Zhang, and Q. Zhang. 2013. Accuracy of genomic prediction for milk production traits in the Chinese Holstein population using a reference population consisting of cows. J. Dairy Sci. 96:5315–5323.

Duchemin, S. I., C. Colombani, A. Legarra, G. Baloche, H. Larroque, J. M. Astruc, F. Barillet, C. Robert-Granie, and E. Manfredi. 2012. Genomic selection in the French Lacaune dairy sheep breed. J. Dairy Sci. 95:2723–2733.

Elzo, M. A., G. C. Lamb, D. D. Johnson, M. G. Thomas, I. Misztal, D. O. Rae, C. A. Martinez, J. G. Wasdin, and J. D. Driver. 2012. Genomic-polygenic evaluation of Angus-Brahman multibreed cattle for feed efficiency and postweaning growth using the Illumina 3K chip. J. Anim. Sci. 90:2488–2497.

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.

Garrick, D. J. 2011. The nature, scope and impact of genomic prediction in beef cattle in the United States. Genet. Sel. Evol. 43:17.

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. 2009. Genomic selection: Prediction of accuracy and maximisation of long term response. Genetica (The Hague) 136:245–257.

Gredler, B., H. Schwarzenbacher, C. Egger-Danner, C. Fuerst, T. Emmerling, and J. Sölkner. 2010. Accuracy of genomic selection in dual purpose Fleckvieh cattle using three types of methods and phenotypes. In: Proc. 9th World Congr. Genet. Appl. Livest. Prod., Leipzig, Germany. www.kongressband.de/wc-galp2010/assets/pdf/0907.pdf. (Accessed 17 October 2013.)

Guillaume, F., S. Fritz, D. Boichard, and T. Druet. 2008. Short communication: Correlations of marker-assisted breeding values with progeny-test breeding values for eight hundred ninety-nine French Holstein bulls. J. Dairy Sci. 91:2520–2522.

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.

Habier, D., R. L. Fernando, and J. C. M. Dekkers. 2007. The impact of genetic relationship information on genome-assisted breeding values. Genetics 177:2389–2397.

Habier, D., R. L. Fernando, and D. J. Garrick. 2013. Genomic BLUP decoded: A look into the black box of genomic prediction. Genetics 194:597–607.

Habier, D., R. L. Fernando, K. Kizilkaya, and D. J. Garrick. 2011. Extension of the Bayesian alphabet for genomic selection. BMC Bioinf. 12:186.

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

Hozé, C., M. N. Fouilloux, E. Venot, F. Guillaume, R. Dassonneville, S. Fritz, V. Ducrocq, F. Phocas, D. Boichard, and P. Croiseau. 2013. High-density marker imputation accuracy in sixteen French cattle breeds. Genet. Sel. Evol. 45:33.

Jimenez-Montero, J. A., O. Gonzalez-Recio, and R. Alenda. 2012. Genotyping strategies for genomic selection in small dairy cattle populations. Animal 6:1216–1224.

Karoui, S., M. J. Carabano, C. Diaz, and A. Legarra. 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genet. Sel. Evol. 44:39.

Kizilkaya, K., R. L. Fernando, and D. J. Garrick. 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. J. Anim. Sci. 88:544–551.

Legarra, A., A. Ricard, and O. Filangi. 2013. GS3-genomic selection, Gibbs sampling, Gauss Seidel and Bayes Cπ. http://snp.toulouse.inra.fr/~alegarra. (Accessed 17 October 2013.)

Liu, Z. T., F. R. Seefried, F. Reinhardt, S. Rensing, G. Thaller, and R. Reents. 2011. Impacts of both reference population size and inclusion of a residual polygenic effect on the accuracy of genomic prediction. Genet. Sel. Evol. 43:19.

Lund, M. S., A. P. W. de Roos, A. G. de Vries, T. Druet, V. Ducrocq, S. Fritz, F. Guillaume, B. Guldbrandtsen, Z. T. Liu, R. Reents, C. Schrooten, F. Seefried, and G. S. Su. 2011. A common reference population from four European Holstein populations increases reliability of genomic predictions. Genet. Sel. Evol. 43:43.

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.

Ostersen, T., O.F. Christensen, M. Henryon, B. Nielsen, G. Su G, and P. Madsen, 2011. Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. Genet. Sel. Evol. 43:38

Pryce, J. E., J. Arias, P. J. Bowman, S. R. Davis, K. A. Macdonald, G. C. Waghorn, W. J. Wales, Y. J. Williams, R. J. Spelman, and B. J. Hayes. 2012. Accuracy of genomic predictions of residual feed intake and 250-day body weight in growing heifers using 625,000 single nucleotide polymorphism markers. J. Dairy Sci. 95:2108–2119.

Rius-Vilarrasa, E., R. F. Brondum, I. Stranden, B. Guldbrandtsen, E. Strandberg, M. S. Lund, and W. F. Fikse. 2012. Influence of model specifications on the reliabilities of genomic prediction in a Swedish-Finnish red breed cattle population. J. Anim. Breed. Genet. 129:369–379.

Saatchi, M., M. C. McClure, S. D. McKay, M. M. Rolf, J. Kim, J. E. Decker, T. M. Taxis, R. H. Chapple, H. R. Ramey, S. L. Northcutt, S. Bauck, B. Woodward, J. C. M. Dekkers, R. L. Fernando, R. D. Schnabel, D. J. Garrick, and J. F. Taylor. 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genet. Sel. Evol. 43:40.

Saatchi, M., J. Ward, and D. J. Garrick. 2013. Accuracies of direct genomic breeding values in Hereford beef cattle using national or international training populations. J. Anim. Sci. 91:1538–1551.

Schaeffer, L. R. 2006. Strategy for applying genome-wide selection in dairy cattle. J. Anim. Breed. Genet. 123:218–223.

Solberg, T. R., A. K. Sonesson, J. A. Woolliams, J. Odegard, and T. H. E. Meuwissen. 2009. Persistence of accuracy of genome-wide breeding values over generations when including a polygenic effect. Genet. Sel. Evol. 41:53.

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

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.

Thomasen, J. R., B. Guldbrandtsen, G. Su, R. F. Brondum, and M. S. Lund. 2012. Reliabilities of genomic estimated breeding values in Danish Jersey. Animal 6:789–796.

Tsuruta, S., I. Misztal, and T. J. Lawlor. 2013. Short communication: Genomic evaluations of final score for US Holsteins benefit from the inclusion of genotypes on cows. J. Dairy Sci. 96:3332–3335.

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