Proceedings, 10th World Congress of Genetics Applied to Livestock Production

Accuracy of Genomic Prediction in French Charolais Cattle Population Using High-density Chip

***T. Tribout*[*], M. Gunia[*,†], R. Saintilan[‡], E. Venot[*], M.N. Fouilloux[§] and F. Phocas[*]**
[*]INRA, UMR1313 GABI, F-78350 Jouy-en-Josas, France, [†]INRA, UMR1388 GenPhySE, F- 31326 Castanet-Tolosan,
[‡]UNCEIA, F-78350 Jouy-en-Josas, France, [§]Institut de l'Elevage - Idele, F-78350, France

ABSTRACT: Genomic breeding values were estimated in the French Charolais beef breed for the direct and maternal genetic effects for field traits routinely recorded at birth and weaning. A total of 2,671 (for direct effect traits) and 872 (for maternal effect traits) animals genotyped for the BovineHD BeadChip or imputed from the BovineSNP50 BeadChip were clustered into training (TS) and validation sets (VS). Their deregressed EBV were used as observations in weighted analyses using a BayesC approach. Accuracies of genomic EBV (GEBV) varied from 0.48 to 0.61 for direct effect traits and from 0.29 to 0.50 for maternal effect traits when the VS animals had close relatives in the TS. The GEBV of VS animals having only more distant relatives in the TS were on average 18% and 51% less accurate for direct and maternal effect traits, respectively.
Keywords: beef cattle; genomic selection

Introduction

Genomic selection (GS) can potentially increase genetic gain in a livestock breeding scheme. Among other strengths, this methodology can improve the accuracy of the breeding value estimates of young selection candidates that do not necessarily have their own performance record or progeny information. Accuracy may be influenced by many factors, including the training set size (Lund et al. (2010)) and the relatedness between the candidates and the training set (Clark et al. (2012)). Pools of genotyped animals have been recently constituted in the main French beef cattle breeds, which could be used as training set making possible to consider the implementation of routine GS. The aim of this study was to estimate accuracy and bias of genomic value prediction in the French Charolais breed under different genetic models, considering close or weak relatedness between the evaluated candidates and the training set.

Materials and Methods

Phenotype data. Five field data traits included in the national pedigree-based BLUP genetic evaluation of the French Charolais breed were considered: birth weight (BW), calving ease score (CE), weaning weight (WW), as well as muscular development (MD) and skeletal development (SD) scores at weaning. Our objective was to compute genomic predictions for the direct genetic effect of animals for these five traits (called dirBW, dirCE, dirWW, dirMD and dirSD in the following) and for the maternal genetic effect of animals for BW, CE and WW (called matBW, matCE and matWW hereafter, respectively). The heritabilities of the 8 traits are given in Table 1.

The response variables used to estimate the SNP effects were deregressed EBV (DEBV). The DEBV were calculated from the pedigree-based BLUP direct and maternal EBV of the genotyped animals and their parents following the method described by Garrick et al. (2009). This method results in DEBV that are free of parent average effects. The DEBV were then used in weighted analyses, to account for heterogeneous variances due to a various amount of progeny records among the genotyped animals.

Genotype data. A total of 2,671 registered French Charolais animals (94% males), born between 1965 and 2012, were genotyped, either with the Bovine SNP50 (50K) BeadChip® or with the BovineHD (777K) BeadChip®.

After usual quality controls, imputation of 50K genotypes to 777K genotypes was performed using BEAGLE software (Browning and Browning (2009)) following the procedure described in Hozé et al. (2013). The numbers of genotyped animals with DEBV varied from 795 to 872 for maternal effects, and from 2,301 to 2,671 for direct effects, and are presented for each trait in Table 1.

Scenarios. Basically, the animals available for the study were clustered into a training set (TS) used to establish prediction equations of the genomic estimated breeding values (GEBV), and into a validation set (VS) on which these equations were applied to estimate the GEBV of its members. The accuracy and prediction bias of the GEBV were respectively estimated in the VS as the weighted correlation between the GEBV and the DEBV, and as the weighted regression coefficient of DEBV on GEBV (RCo).

Two different clustering rules were applied in order to evaluate the accuracy of GEBV according to the relatedness between the candidates and the TS.

Weak Relatedness scenario (WR). In a first case, the VS for the direct and maternal traits were respectively constituted by the 333 and 174 animals that had a DEBV for all the direct or maternal traits and that had no genotyped parent and no genotyped offspring in the TS.

Close Relatedness scenario (CR). In a second case, the VS for the direct and maternal traits were respectively made up of 330 and 193 animals with a DEBV for all the direct or maternal traits, and that had no genotyped

Table 1. Heritability (h^2), size of training and validation sets (n), and average reliability (rel) of deregressed EBV (DEBV)[1] for all considered traits in the scenarios of close and weak relatedness between validation and training sets.

		Weak relatedness scenario				Close relatedness scenario				
		Training set		Validation set		Training set		Validation set		Total
Trait	h^2	n	rel	n	rel	n	rel	n	rel	n
Maternal effect traits										
Birth Weight	0.06	698	0.55	174	0.53	679	0.57	193	0.47	872
Calving ease	0.05	681	0.54	174	0.50	662	0.56	193	0.43	855
Weaning weight	0.07	621	0.53	174	0.45	602	0.57	193	0.35	795
Direct effect traits										
Birth Weight	0.36	2,338	0.52	333	0.65	2,341	0.49	330	0.84	2,671
Calving ease	0.09	2,322	0.30	333	0.44	2,325	0.27	330	0.64	2,655
Weaning weight	0.22	2,038	0.42	333	0.53	2,041	0.40	330	0.68	2,371
Muscular development	0.30	1,968	0.49	333	0.59	1,971	0.47	330	0.73	2,301
Skeletal development	0.27	1,968	0.47	333	0.57	1,971	0.45	330	0.71	2,301

[1]Deregressed EBV corrected from the parent average contribution

offspring but at least one parent and one grandparent in the TS.

In both cases, the TS were constituted by all available animals not being part of the VS. The average reliability of the DEBV in the TS and VS are presented in Table 1.

Statistical models. Genomic prediction equations were derived using a BayesC approach (Habier et al. (2011)). The π fraction of SNP with a non null effect within a replicate was fixed at a value of 0.001 (approximately 700 markers among 708,771). Analyses were carried using GS3 software (Legarra et al. (2013)). For each analysis, 50,000 iterations were run with a burn-in of 10,000 and a thin of 10.

General Model. For each trait, the following model was fit to the response variable DEBV for TS: $\mathbf{y} = \mathbf{1}\mu + \mathbf{Ma} + \mathbf{e}$ (1)

where $\mathbf{y}$ is the vector of DEBV of TS animals, $\mathbf{1}$ is a vector of 1, μ is the overall mean, $\boldsymbol{M}$ is the incidence matrix for SNP genotypes indicating the number of copies of a given marker allele carried by the individuals, $\mathbf{a}$ is the vector of marker effects, and $\mathbf{e}$ is the vector of residual effects. This model will be referred as GM in the following.

Once the markers effects were estimated, the GEBV of any individual in the VS was predicted as $GEBV_i = \sum_{j=1}^{J} M_{ij}\,\hat{a}_j$,

where $\hat{a}_j$ is the estimated effect of SNP *j*, M_{ij} is the genotype of individual *i* for SNP *j*, and *J* is the total number of markers.

Alternative models. The GM model assumes that the entire genetic variance (σ_g^2) is captured by the markers. This is however unlikely, mainly because of the limited size of the TS and the partial LD between SNP and causal genes. Therefore, two alternative models were tested:

The first alternative consisted to simply include the fraction of genetic variability assumed not captured by the markers in the residual part of the model, without trying to predict polygenic EBV. The fitted model was therefore similar to model (1), but the computation of the DEBV weights differed, as described in Garrick et al. (2009).

The second alternative consisted to explicitly include a polygenic fraction in the model, as: $\mathbf{y} = \mathbf{1}\mu + \mathbf{Ma} + \mathbf{Zu} + \mathbf{e}$ where $\mathbf{u}$ is a vector of polygenic effects and $\mathbf{Z}$ is an incidence matrix for the polygenic effects. In that case, the GEBV value for the VS individual was the sum of the predicted effects of the SNP over all the markers and of the polygenic EBV: $\widehat{GEBV}\imath = \sum_{j=1}^{J} M_{ij}\,\hat{a}_j + \hat{u}_\imath$

These alternative models were applied assuming 20% and 40% of σ_g^2 was not captured by markers.

Pedigree-based BLUP. Finally, the pedigree-based BLUP (pBLUP) EBV of the VS animals were estimated as a reference method, using the DEBV of the TS and the pedigree relationships between TS and VS.

Results and Discussion

Accuracy of GEBV.

General Model. The GEBV accuracy obtained with the GM model for direct effect traits varied from 0.38 to 0.54 under the WR scenario, and from 0.48 to 0.60 under the CR scenario (Table 2). For these traits, genomic evaluations resulted in more accurate EBV than pBLUP evaluations based only on the TS data and pedigree relationships.

The GEBV for maternal effect traits were less accurate than for direct effect traits, ranging from 0.12 to 0.27 and from 0.29 to 0.50 in the WR and CR scenarios, respectively. The GEBV accuracy remained lower for maternal than for direct effect traits even after adjustments to account for DEBV reliability differences in the VS between traits and

Table 2. Accuracy[1] of genomic EBV and pedigree-based BLUP EBV in the validation set according to the statistical model used, in the scenarios of close and weak relatedness between validation and training sets.

	Weak relatedness scenario						Close relatedness scenario					
	Polygenic fraction of the genetic variance						Polygenic fraction of the genetic variance					
Trait	0%[2]	20% in the model	40% in the model	20% in the residual	40% in the residual	pBLUP[3]	0%[2]	20% in the model	40% in the model	20% in the residual	40% in the residual	pBLUP[3]
Maternal effect traits												
Birth Weight	0.12	0.11	0.11	0.08	0.07	0.07	0.29	0.29	0.29	0.32	0.32	0.28
Calving ease	0.18	0.18	0.20	0.20	0.19	0.19	0.39	0.40	0.41	0.37	0.35	0.43
Weaning weight	0.27	0.30	0.33	0.23	0.22	0.27	0.50	0.50	0.50	0.49	0.48	0.50
Mean	*0.19*	*0.20*	*0.21*	*0.17*	*0.16*	*0.18*	*0.39*	*0.40*	*0.40*	*0.39*	*0.38*	*0.40*
Direct effect traits												
Birth Weight	0.44	0.44	0.42	0.41	0.39	0.20	0.60	0.60	0.59	0.59	0.57	0.50
Calving ease	0.47	0.49	0.49	0.40	0.39	0.33	0.48	0.49	0.48	0.54	0.53	0.39
Weaning weight	0.38	0.40	0.42	0.41	0.40	0.24	0.53	0.54	0.55	0.50	0.48	0.50
Muscular development	0.48	0.49	0.48	0.48	0.47	0.27	0.57	0.58	0.58	0.59	0.58	0.41
Skeletal development	0.54	0.56	0.57	0.53	0.52	0.44	0.59	0.60	0.60	0.61	0.61	0.56
Mean	*0.46*	*0.47*	*0.48*	*0.45*	*0.43*	*0.30*	*0.56*	*0.56*	*0.56*	*0.57*	*0.55*	*0.47*

[1]Accuracy is measured by the weighted Pearson's correlation in the validation sets between the DEBV and the genomic or polygenic EBV
[2]GM model with 100% of the genetic variance assumed captured by SNP
[3]Pedigree-based BLUP evaluation based on the training set data and on pedigree relationships

scenarios (not presented). This result can be explained by the smaller size of the TS for maternal than for direct effect traits, and is consistent with several studies showing the importance of TS size on the GEBV accuracy (e.g. Lund et al. (2011)). Genomic evaluations and pBLUP evaluations based only on the TS data resulted in similar accuracies for maternal effect traits, probably because of the limited size of the TS which did not allow reliable estimates of SNP effects.

Comparing our results with those of other authors is difficult, mainly because of differences in the size and reliability of the TS, as well as in the genetic structure of the TS and VS across the studies. Keeping these limitations in mind, the values of accuracies estimated here for dirBW, dirCE and matCE are consistent with those estimated by Saatchi et al. (2011) in American Angus cattle. Concerning dirWW and matWW, our results are slightly higher than the values reported by these authors.

Table 2 shows that the GEBV accuracy was higher in the CR scenario than in the WR scenario (+20% and +100% for direct and maternal effect traits, respectively). This illustrates the finding of Clark et al. (2012) that the accuracy of GS increases when the candidates have close relatives in the TS. A large part of the genotyped animals used in the current study was constituted by AI bulls. In the French Charolais population, only 13% of the calves are bred by AI sires. There is therefore a risk that a majority of candidates wouldn't have any close relative in the TS if GS was implemented, resulting in low efficiency of selection for these animals. Here, a large proportion of the VS animals in the WR scenario had one or more grand-parents or half sibs in the TS. This would probably not be the case in practice for the 71% of calves that are bred by NS sires and NS paternal grandsires. Therefore, their GEBV accuracy could actually be even lower than estimated in the WR case. This emphasizes the need to genotype as many breeding animals with their own phenotype or phenotyped offspring as possible, in order to constitute the most informative TS.

Alternative models. Overall, including a polygenic term in the model explaining 20 or 40% of σ_g^2 had little effect on the GEBV accuracy. A weakly favorable effect was observed for matCE (CR and WR cases), dirWW, dirSD and matWW (WR case), but it tended to reduce the GEBV accuracy for dirBW and matBW under the WR scenario. The model deporting a fraction of σ_g^2 in the residual term had no effect on the accuracy for dirDM, dirDS (CR and WR cases) and matWW (CR case). A slightly favorable effect was observed for dirCE and matBW in the CR scenario, as well as for dirWW and matCE in the WR scenario, but the accuracy decreased in all other cases. However, given the limited size of the VS, most of these differences are probably not significant.

Prediction bias. The GEBV tended to be biased for most of the traits under the GM model. On average, RCo was lower than one for maternal effect traits (0.83 and 0.77 in the CR and WR cases, respectively) and direct effect traits in the CR case (0.86), but was equal to 1.11 for the direct effects traits under the WR scenario. Adding a polygenic component in the residual or explicitly in the model increased RCo, and had therefore a favorable effect on the prediction bias, except for direct effects traits under the WR scenario for which the bias worsened.

Conclusion

Though the accuracies presented here are moderate, these results are promising and show that implementing GS in the French Charolais breed for birth and weaning traits is possible. The pool of Charolais animals available to establish the

genomic prediction equations is currently small, in particular for maternal effect traits, but it would progressively enlarge in a routine genomic evaluation as new phenotyped candidates would be genotyped. Our results show the need to genotype as many breeding animals as possible, in order to allow a maximum of candidates to have close relatives in the TS, and thus to have more accurate GEBV. However, thousands of Charolais bulls and cows that do not belong to the nucleus population have no EBV in the current system. It would thus be very useful for breeders to have GEBV available to select these animals, even if their accuracy is moderate.

Acknowledgments

GEMBAL project (ANR-10-GENM-0014) is funded by ANR, ApisGene, Races de France and INRA. The 50K genotypes originated from others ANR-ApisGene projects and from French cattle breeding companies' activity.

Literature Cited

Browning, B.L., Browning, S.R. (2009). *Am. J. Hum. Genet.* 84:210-223.

Clark, S.A., Hickey, J.M., Daetwyler, H.D., van der Werf, J.H.J. (2012). *Genet. Sel. Evol.* 44:4.

Garrick, D.J., Taylor, J.F.T., Fernando, R.L.F. (2009). *Genet. Sel. Evol.* 41:55.

Habier, D., Fernando, R.F., Kizilkaya, K., Garrick, D.J. (2011). *BMC Bioinformatics* 12:186.

Hozé, C., Fouilloux, M.N., Vénot, E. et al. (2013). *Genet. Sel. Evol.* 45:33.

Legarra, A., Ricard, A., Filangi, O. (2013). http://snp.toulouse.inra.fr/~alegarra.

Lund, M.S., de Roos, A.P.W., de Vries, A.G. et al. (2011). *Genet. Sel. Evol.* 43:43.

Saatchi, M., McClure, M.C., McKay, S.D. et al. (2011). *Genet. Sel. Evol.* 43:40.