Proc. Assoc. Advmt. Anim. Breed. Genet. 20:348-351

USE OF HIGH DENSITY GENOTYPING AND TRAIT-DEPENDENT METHODS IN GENOME-ASSISTED EVALUATIONS

J.A. Jiménez-Montero[1], D. Gianola[2,3], K. Weigel[2], R. Alenda[1] and O. González-Recio[4]

[1]Dpt Producción Animal. ETSI Agrónomos. Universidad Politécnica de Madrid. 28040 Madrid, Spain
[2]Department of Dairy Science, University of Wisconsin-Madison, 1675 Observatory Dr., WI 53076,USA
[3]Department of Animal Sciences, University of Wisconsin-Madison, 1675 Observatory Dr., WI 53076,USA
[4]Dpto Mejora Genética Animal. INIA. Ctra. La Coruña km 7,5. 28040 Madrid

Presenting author: gonzalez.oscar@inia.es

SUMMARY

This study presents the predictive ability in genome-assisted evaluations using different density arrays and two different statistical methods. Predictive ability from the Genomic BLUP (**G-BLUP**) using 1632 progeny tested sires genotypes with the BovineSNP50v2 BeadChip (**50K**) were considered as benchmark. First, genotypes from the BovineLD BeadChip (**LD6K**) were imputed to a BovineSNP50v2 BeadChip (50K). Second, genotypes were imputed to the BovineHD BeadChip (**HD**). The Random Boosting (**R-Boost**) was evaluated as an alternative method to G-BLUP. Four traits were analyzed: milk yield (**MY**), fat percentage (**FP**), somatic cell count (**SCC**) and days open (**DO**).

In general, R-Boost and G-BLUP showed similar results with overlapping confidence interval. Low density genotypes imputed to 50K achieved a similar predictive ability than native 50K genotypes. However, an increase in Pearson correlation and lower predictive mean square error were found across traits when genotypes were imputed to HD. The larger improvements were found for DO when using imputed HD genotypes (up to 0.06 greater Pearson correlation units) and for FP using R-Boost (up to 0.20 greater Pearson correlation units).

These results showed that the predictive ability of certain traits may be improved either imputing genotypes to HD or utilizing a method that get adapted to the genetic architecture of trait.

INTRODUCTION

The next key objective in genomic selection programs is to translate the huge, and increasing, amount of genomic information in a useful tool to breeders (Pryce and Daetwyler 2012). Low density SNP panels and posterior imputation is a promising way to reduce genotyping costs while maintaining a large predictive reliability. There is a need to integrate different density SNP panels in genomic breeding programs. Further, there may be an interaction between the density of the original genotype and the statistical method for DGV prediction, and these may be trait dependent as well. It is known that methods based on marker regression have better predictive ability than methods based on genomic relationship matrices in traits that are regulated by major genes. Higher-density arrays are expected to capture a larger amount of genetic variance because LD between markers and causal mutations is supposed to be higher. However, previous studies have shown only a slight increase in predictive accuracy using arrays of up to 700K SNPs (VanRaden *et al*. 2013).

The objective of this study was to compare imputation accuracy, predictive ability, and selection efficiency for selection candidates genotyped at different densities using the Random Boosting (R-Boost) and G-BLUP algorithms.

MATERIAL AND METHODS

Genotypes and phenotypes. A total of 2658 genotyped bulls were used in this study, using the BovineSNP50.v2 Beadchip for 2226 bulls and the BovineSNP50.v1 Beadchip (Illumina Inc,) for 240 bulls. These 2658 bulls build up the 50K Spanish Holstein population (50K). Additionally, 192 were genotyped using the BovineHD BeadChip (HD). Editing and filtering processes of genotypes led to 39,714 and 540,501 SNPs for the 50K and HD evaluations, respectively.

A total of 1632 progeny tested bulls born before 2006 were used as the reference set (1412 for DO), labeled as **TRAIN50K**. Bulls born between 2006 and 2010 were used as the validation sets (382 for MY, FP and SCC and 216 for DO), labeled as **TEST50K**.

Four complex traits were examined: milk yield (**MY**), fat percentage (**FP**), somatic cell count (**SCC**) and days open (**DO**). These traits were selected to show differences regarding heritability of the trait and amount of phenotypic information available. Deregressed MACE progeny proofs (**DRP**) from January 2009 Interbull evaluation (Jairath *et al*. 1998) were used to estimate marker coefficients in the reference set.

Imputation. Low density genotypes in the testing set were created *in silico*, masking SNPs included in the 50K assay that were not included in the Bovine LD (LD6K) (Illumina Inc.) assay. Thereafter, phased haplotypes from 1632 animals in TRAIN50K were used as reference set for imputing the LD6K validation set using Beagle (Browning and Browning 2009). The outcomes were referred as 6K50K. Then, imputation from 50K (6K50K, TEST50K and TRAIN50K) to HD (6KHD, 50KHD and TRAIN50KHD) was implemented using the original HD population as reference.

Genomic evaluation models. Two different genomic evaluation models were used: Random boosting (R-Boost) (Gonzalez-Recio *et al*., 2013) and G-BLUP (VanRaden, 2008).

Predictive ability. *Accuracy and predictive MSE.* The prediction accuracy of genomic evaluations was computed as the Pearson correlation between the predicted DGVs and the December 2011 DRPs. The PMSE of predictions was also estimated. Means and confidence intervals were estimated using bootstrapping for each evaluation output (Efron, 1986), although these results are not shown in this work.

RESULTS AND DISCUSSION

Imputation Performance The allele error rate at imputing genotypes from customized LD6K to 50K was 1.3%. Those results are in accordance to previous studies using similar population sizes (Zhang and Druet, 2010; Berry and Kearney, 2011; Dassonneville *et al*. 2012). The LD6K array is an important tool for candidates preselection and genotyping females. Allele error rate of imputation from 50K to HD was 0.9% when a small reference population of 192 bulls genotyped in HD was used.

Predictive ability Pearson correlations obtained from the two methods are shown in Table 1. Both methods resulted in similar accuracy, although R-Boost was the preferred method for FP and G-BLUP for MY and DO, while for SCC results were case dependant.

Table 1. Accuracy for the genomic estimation of two evaluation methods for four traits of economic interest in dairy cattle after imputation from 6K and 50K to 50K and HD.

Trait	Method	6K50K	TEST50K	6KHD	50KHD
Milk yield (MY)	G-BLUP	**0.59**	**0.59**	0.54	0.55
	R-Boost	0.55	0.57	0.54	0.54
Fat percentage (FP)	G-BLUP	0.60	0.60	0.55	0.55
	R-Boost	0.78	0.78	0.79	**0.80**
Somatic cell count (SCC)	G-BLUP	0.49	0.48	**0.50**	0.47
	R-Boost	0.45	0.46	**0.50**	0.49
Days open (DO)	G-BLUP	0.29	0.19	**0.32**	0.31
	R-Boost	0.19	0.22	0.20	0.28

In bold: The preferred method within trait and set criteria

Prediction accuracy slightly increased for all traits after imputation to HD, excepting for MY. This increment was more relevant (up to 6 points in accuracy) for DO, which was the analyzed trait with smaller heritability. These results were in accordance with results previously reported for other Holstein populations, where estimates from HD were slightly better than those from 50K (Erbe *et al*. 2012, VanRaden *et al*. 2013).

Confidence intervals estimated by bootstrapping showed that distributions regarding prediction accuracy widely overlapped across methods and sets for MY and SCC (results not shown). However, R-Boost estimates were more accurate than G-BLUP for FP. As expected, large bootstrapped confidence intervals were found for DO (results not shown).

The MSE of prediction showed notable differences between evaluation methods (Table 2). R-Boost showed smaller PMSE in all four traits (12%, 54%, 12%, and 5% smaller for MY, FP, SCC, and DO, respectively). Note that R-Boost aims to minimize the MSE, as this is assumed as the loss function. Those results, and the aforementioned accuracies were in agreement with (Jiménez-Montero *et al*., 2013), although bootstrap analyses showed no significant differences. As before, lower PMSE were obtained after imputation to HD, excepting for MY.

CONCLUSIONS

Imputation using Beagle software was efficient for the reconstruction of 50K genotypes from low-density chips. Genomic evaluation methods (R-Boost and G-BLUP) resulted in similar prediction ability for the traits and genotypes included in this study. R-Boost showed clearly better performance for FP, and in terms of PMSE for all traits. However, no clear differences were found in terms of accuracy.

In general, some improvement in the predictive accuracy was obtained after imputation to HD. Genetic and genomic evaluation units should consider using different methods regarding the trait evaluated, and imputation to HD might be interesting to increase the predictive ability of some traits, especially those of low heritability or those regulated by major genes (e.g. FP, SCC, DO).

Table 2. Mean Squared Errors in the validation set regarding genomic evaluation method and imputation strategy (from 6K and 50K to 50K and HD) for milk yield, fat percentage, somatic cell count and days open.

Trait	Method	6K50K	TEST50K	6KHD	50KHD
Milk yield (MY)	G-BLUP	255	258	276	278
	R-Boost	236	**229**	241	240
Fat percentage (FP)	G-BLUP	0.044	0.044	0.048	0.047
	R-Boost	0.030	0.030	0.028	**0.027**
Somatic cell count (SCC)	G-BLUP	155.1	154.5	152.0	143.3
	R-Boost	138.3	137.4	**131.7**	133.4
Days open (DO)	G-BLUP	548.6	636.3	530.2	535.1
	R-Boost	541.9	523.4	546.4	**519.6**

In bold: The preferred method within trait and set criteria

ACKNOWLEDGMENTS

The authors acknowledge funds from the project CDTI-P080250866 UPM and the agreement INIA-CC10-046, to CONAFE, EUROGENOMICS consortium, ASCOL, ABEREKIN, XENETICA FONTAO and GENETICAL for providing biological samples and phenotypes used in this study and to "Dirección General de Producciones y Mercados Agrarios" ,"Laboratorio Central de Veterinaria del Ministerio de Agricultura, Alimentación y Medio Ambiente" for support of genotyping process and specially Dr. Dassonneville for helpful suggestions and comments.

REFERENCES

Berry D.P. and Kearney J.F. (2011) *Animal* **5:**1162

Browning B.L. and Browning S.R. (2009) *The American Journal of Human Genetics* **84:** 210

Dassonneville R., Fritz S., Ducrocq V. and Boichard, D. (2012) *Journal of Dairy Science* **95**: 4136

Efron, B., (1986) Statist. Sci. **1:** 54

Erbe M., Hayes B.J., Matukumalli L.K., Goswami S., Bowman P.J., Reich C.M., Mason B.A., and Goddard M.E. (2012) *Journal of Dairy Science* **95**: 4114

González-Recio O., Jiménez-Montero J.A. and Alenda R. (2013) *Journal of Dairy Science* **96**: 614

Jairath L., Dekkers J.C.M., Schaeffer L.R., Liu Z., Burnside E.B. and Kolstad B. (1998) *Journal of Dairy Science* **81**: 550

Jiménez-Montero J.A., González-Recio O. and Alenda R. (2013) *Journal of Dairy Science* **96**: 625

Mäntysaari E., Liu Z. and VanRaden P. (2010) *Interbull Bull* **41**: 17

Pryce J.E. and Daetwyler H.D. (2012) *Anim. Prod. Sci*. **52**: 107

VanRaden P.M., (2008) *Journal of Dairy Science* **91**: 4414

VanRaden P.M., Null D.J., Sargolzaei M., Wiggans G.R., Tooker M.E., Cole J.B., Sonstegard T.S., Connor E.E., Winters M., van Kaam J.B.C.H.M., Valentini A., Van Doormaal B.J., Faust M.A. and Doak G.A. (2013) *Journal of Dairy Science* **96**: 668

Zhang Z. and Druet T. (2010) *Journal of Dairy Science* **93**: 5487