Genotyping Strategies for Genomic Selection in Dairy Cattle

J.A. Jiménez-Montero[*], O. Gonzalez-Recio[†], R. Alenda[*]

Introduction

Genomic selection is one of the most promising tools to increase genetic gain rate appeared in the last decades. It has been focused on predicting sires' PTA or DYD due to the higher accuracy of proven bulls' EBVs and the impact of sires in the breeding programs. Nonetheless, there are some facts that make hypothesizing that genomic selection may be of interest in the female population. For instance, economically important traits are measured in the female population, it represents the largest proportion of the Holstein population, dominant and epistatic effects may be captured and exploited on it and the association between cow's genotype and 'adjusted phenotype' is expected to be stronger than that between sire's genotype and his progeny average. As current genotyping cost is still economically unfeasible for commercial farms, maximizing genotyping investment is still a great challenge in genomic selection to be used in the commercial population.

Several genotyping strategies may be considered. For instances, a reduced number of informative SNPs may be chosen from a reference population, usually sires (Van Raden *et al* 2009; Weigel *et al.,* 2009). The use of imputation algorithms for obtaining unobserved SNPs genotypes using low dense panels as described in Weigel *et al.* (accepted). Selective phenotyping and genotyping of most informative animals (Sen *et al.,* 2009; Spangler *et al.,* 2008) is also an interesting strategy that may be compatible with the previous ones. This study focuses on this latter.

The objective of the present study was to evaluate the most informative case in a selective genotyping strategy to increase the predictive accuracy of EBV in future generations.

Material and methods

Simulation. QMSim software (Sargolzaei and Schenkel, 2009) was run according to parameters showed in Table 1. First, 1040 historical generations were generated to produce a realistic level of LD. Then, 20,000 females and 300 males were selected as founders, followed by 10 generations of selection. Animals in the generations 11 to 14 were used as reference population (40,195), while whole generation 15 was genotyped as test set. These simulations try to mimic a Spanish dairy population.

Selective genotyping strategies

The 2%, 5% and 10% of the reference population were selected as training set according to different strategies for both 0.25 and 0.10 heritability traits, as described next:

1. At random (**RND**).- Females drawn randomly from the whole reference population.

[*] Departamento de Producción Animal, E.T.S. Agrónomos-Universidad Politécnica,.Ciudad Universitaria s/n, 28040 Madrid, Spain

[†]Departamento de Mejora Genética Animal, INIA, Carretera La Coruña km 7, 28040 Madrid, Spain

2. Divergent phenotypic values (**DPH**).- Equal number of females in the α and (1-α) percentiles of the 'adjusted' phenotypic distribution.
3. Divergent EBV values (**DBV**).- Females with their breeding values in the α and (1-α) percentiles.
4. Highest phenotypic values (**TopPH**).- Top ranking cows for 'adjusted' phenotypic values
5. Highest EBV values (**TopBV**).- Top ranking cows for breeding values.
6. Divergent family EBV (**DFM**).- Half sibs females sired only by the best and worse bulls in EBV.

As benchmark, all contemporary sires (996) from G-11 to G-14 were also genotyped in accordance with the most common current strategy (**SiresDYD**). It must be pointed out that the phenotype used in this study may be interpreted as an adjusted phenotype by environmental effects, resulting in a combination of cow's genetic value plus some residual.

Table 1: Parameters used in the simulations

Heritability	0.25	0.10
Phenotypic variance	1.0	1.0
Historical population generation	1040	1040
Founders (males/females)	300 / 20,000	300 / 20,000
Mutation rate	2.5e-5	2.5e-5
Selection generations	15	15
SNP Platform	10,000	10,000

Genomic evaluation model.
Bayesian Lasso (de los Campos *et al.,* 2009) was used to estimate SNP coefficients on the genotyped population (according to the strategies above) and predicting the corresponding genomic breeding values (GBV) in generation 15. Phenotypes were used as dependent variables in all strategies, except for SiresDYD, in which DYD were used. Pearson correlations between predicted GBV and true breeding values (TBV) were calculated in generation 15, and confident intervals were obtained using Boostraping (Efron, 1979).

Results and discussion

Table 2 shows some parameters describing the simulated population. A linkage disequilibrium of 0.33 (±0.02), measured as r^2 (Hill and Robertson, 1968), per chromosome in the last generation (test set) was obtained. This value was similar to that reported by Sargolzaei *et al.* (2008) in Holstein Cattle in North America using a 10k SNP platform, as was simulated in this study. An average inbreeding coefficient of 0.04 and 0.07 was achieved in the last generation in each of the scenarios of medium or low heritability, respectively.

Table 2: Genetic parameters of interest in the test population

Parameters	h^2=0.25	h^2=0.10
Inbreeding	0.04±0.01	0.07±0.03
LD (r^{2*})	0.33±0.02	0.33±0.02
Pedigree Index accuracy Generation 15	0.39±0.01	0.36±0.01

***Average Linkage Disequilibrium between adjacent SNP in the 30 Chromosomes**

Predictive accuracy, measured through Pearson correlations between predicted GBV and TBV, are shown in Table 3 for both medium and low heritability scenarios. The accuracies of the SiresDYD strategy were 0.44 and 0.71 for the low and medium heritability traits, respectively. As expected, all strategies on the female population increased the predictive accuracy as larger was the proportion of genotyped animals (Figure 1). For the medium heritability trait there was a threshold accuracy around 0.8. Divergent predictors strategies, DPH (0.46 and 0.73) and DBV (0.49 and 0.73), showed better predictive accuracy than DYD at 2% of population genotyped with h^2=0.25. DPH achieved the largest predictive accuracy (0.66 and 0.81) in both traits.

Table 3: Accuracy[1] of genomic breeding values

h^2	% Genotyped Reference Population	Divergent Values			Top Values		Random
		Phen.	EBV	Fam. EBV	Phe.	EBV	RND
0.25	2%	0.73	0.73	0.55	0.19	-0.18	0.44
	5%	0.79	0.75	0.69	0.37	-0.05	0.59
	10%	0.81	0.76	0.73	0.42	0.08	0.71
0.1	2%	0.46	0.49	0.34	0.00	-0.05	0.14
	5%	0.59	0.52	0.42	0.06	-0.09	0.28
	10%	0.66	0.53	0.52	0.14	-0.02	0.37

[1]All standard deviations ranged between ±0.003 and ±0.007

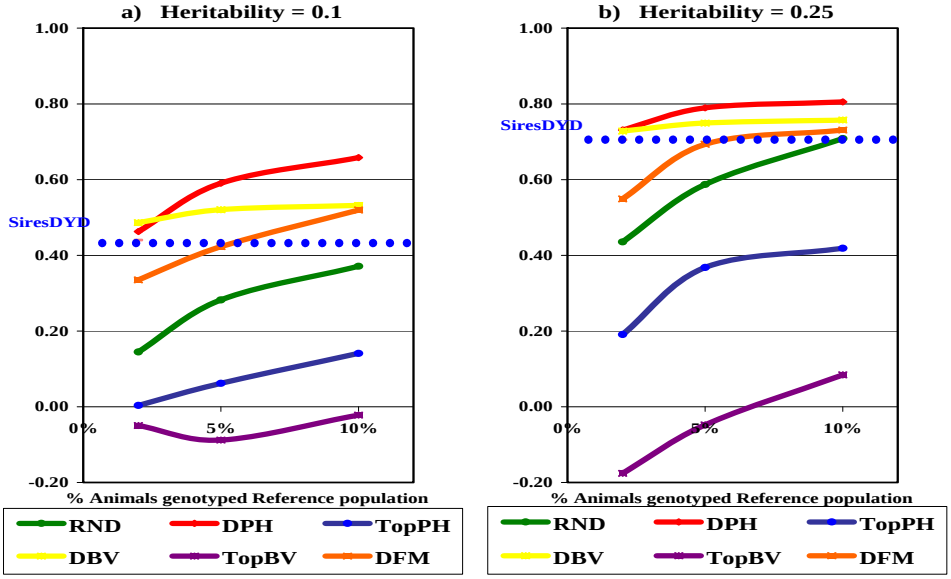

Figure 1: Estimated accuracies for GBV in generation 15, when 2%, 5% and 10% of females in the training population (G 11 – 14) was being genotyped using different strategies. Traits heritability a) 0.10 and b) 0.25

The selective genotyping strategies of TopPH (0.00 and 0.19) and TopBV (-0.05 and -0.18) achieved a lowered predictive accuracy regarding RND (0.14 and 0.44) strategy for 2% of population genotyped (for low and medium heritability traits, respectively). Genotyping top EBV animals for their use as a reference population in the prediction of GBV achieved the poorest accuracy.

DFM achieved an intermediate predictive accuracy (0.34 and 0.55) between individual divergent (DPH and DBV) and RND strategies, and similar to that obtained with SiresDYD for a medium heritability trait and over performing it in a low heritability case at increasing the number of animals in the reference population

Conclusion

Predictive accuracy of GBV depends on the amount of animals genotyped and the selective genotyping strategy used. Divergent genotyping strategies in females could be worth to increase the genomic program efficiency together with the current male genotyping strategies. Nonetheless, some bias was detected in all of these strategies mainly for Top strategies and SiresDYD that deserve further researching.
Future research should focus on the male and female genotyping performance, as well as incorporating other strategies to reduce costs. Furthermore a complete economical research is needed in order to find optimal reference population to be genotyped.

References

Efron, B. (1979). *The Annals of Statistics,* 7(1):1–26.

De los Campos, G., Naya, H., Gianola, D. *et al.* (2009). *Genetics,* 182:375–385.

Hill, W.G., and Robertson, A. (1968). *Theor. Appl. Genet.,* 38:226–231.

R Development Core Team (2009). URL http://www.R-project.org.

Sargolzaei, M., Schenkel, F.S., Jansen, G.B. *et al.* (2008). *J. Dairy Sci.,* 91:2106–2117.

Sargolzaei, M., and Schenkel, F.S. (2009). *Bioinformatics,* 25:680–681.

Sen, S., Johannes, F., and Broman, K.W. (2009). *Genetics*, 181:1613–1626.

Spangler, M.L., Sapp, J.K., Bertrand, M.D. *et al.* (2008). *J. of Anim. Sci.,* 86:2471–2479.

Van Raden, P.M., Van Tassell, C.P., Wiggans, G.R. *et al.* (2009). *J. Dairy Sci.,* 92:16–24.

Weigel, K. A., de los Campos, G., González-Recio, O. *et al.* (2009). *J. Dairy Sci.,* 92:5248–5257.

Weigel, K.A., Van Tassell, C.P., O´Connel, J.R. *et al.* (2010). *Submitted.*