*Animal* (2018), **12:7**, pp 1350–1357 © The Animal Consortium 2017
doi:10.1017/S175173111700283X

animal

# Evaluation of the potential use of a meta-population for genomic selection in autochthonous beef cattle populations

E. F. Mouresan[1], J. J. Cañas-Álvarez[2], A. González-Rodríguez[1], S. Munilla[1,3], J. Altarriba[1,4], C. Díaz[5], J. A. Baró[6], A. Molina[7], J. Piedrafita[2] and L. Varona[1,4†]

*[1]Departamento de Anatomía, Embriología y Genética, Universidad de Zaragoza, 50013 Zaragoza, Spain; [2]Grup de Recerca en Remugants, Departament de Ciència Animal 6i dels Aliments, Universitat Autònoma de Barcelona, 08193 Bellaterra, Barcelona, Spain; [3]Departamento de Producción Animal, Facultad de Agronomía, Universidad de Buenos Aires, 1417 CABA, Argentina; [4]Instituto Agroalimentario de Aragón (IA2), 50013 Zaragoza, Spain; [5]Departamento de Mejora Genética Animal, INIA, 28040 Madrid, Spain; [6]Departamento de Ciencias Agroforestales, Universidad de Valladolid, 34004 Palencia, Spain; [7]MERAGEM, Universidad de Córdoba, 14071 Córdoba, Spain*

(Received 25 April 2017; Accepted 11 August 2017; First published online 2 November 2017)

*This study investigated the potential application of genomic selection under a multi-breed scheme in the Spanish autochthonous beef cattle populations using a simulation study that replicates the structure of linkage disequilibrium obtained from a sample of 25 triplets of sire/dam/offspring per population and using the* BovineHD Beadchip*. Purebred and combined reference sets were used for the genomic evaluation and several scenarios of different genetic architecture of the trait were investigated. The single-breed evaluations yielded the highest within-breed accuracies. Across breed accuracies were found low but positive on average confirming the genetic connectedness between the populations. If the same genotyping effort is split in several populations, the accuracies were lower when compared with single-breed evaluation, but showed a small advantage over small-sized purebred reference sets over the accuracies of subsequent generations. Besides, the genetic architecture of the trait did not show any relevant effect on the accuracy with the exception of rare variants, which yielded slightly lower results and higher loss of predictive ability over the generations.*

**Keywords:** beef cattle, genomic selection, across population, accuracy, meta-population

## Implications

The results of this study indicate that the use of a meta-population within the scope of the Spanish autochthonous beef cattle populations may provide an increase of accuracy for populations with small size or limited economic resources for genotyping. Besides, the advantage of admixed populations seems to be greater when predicting individuals more distant from the training set because the across breed and multi-breed genomic predictions are based more on the LD between markers and quantitative trait loci (QTL) than on family relationships.

## Introduction

The advances in the area of molecular genetics have allowed the development of SNP chips that provide genomic information throughout the genome (Gunderson, *et al*., 2005). Along with the molecular advances, new statistical methods have been developed with the purpose of predicting the breeding values of candidates to selection using genomic information (Meuwissen *et al*., 2001). The potential applications of these methods have been tested through simulation (Meuwissen *et al*., 2001) and in different species such as mice (Legarra *et al*., 2008), aquaculture (Vallejo *et al*., 2017), poultry (Heidaritabar *et al*., 2016) and pigs (Tussell *et al*., 2016).

In cattle, genomic selection is a reality in large (Hayes *et al*., 2009) and medium-sized dairy cattle populations (Reiner-Benaim *et al*., 2017). In beef cattle, it has been implement in some medium or large populations (Lourenco *et al*., 2015; Silva *et al*., 2016). Nevertheless, the full application of genomic selection methods in the beef cattle industry is somewhat questionable. The main drawbacks are the limited census of some beef populations, the great variability of the production systems, the narrow use of artificial insemination and the low quality of phenotypic recording (Berry *et al*., 2016).

In simulation scenarios, several authors (De Roos *et al*., 2009; Kizilkaya *et al*., 2010; Toosi *et al*., 2010) have shown how it is feasible to increase the precision of the genomic predictions by using phenotypic and genomic information provided by several populations. Their results indicate that

† E-mail: lvarona@unizar.es

the use of a combined population is more helpful when the populations involved have diverged for a small number of generations, for populations of reduced size, and for traits of low heritability if high-density genotypes are available. With real data, some studies have obtained promising results (Weber *et al.*, 2012), whereas some others reported almost no advantage from a multi-population genomic evaluation (Karoui *et al.*, 2012; Bolormaa *et al.*, 2013; Chen *et al.*, 2013; Kachman *et al.*, 2013). Though it seems that its potential use should be studied in each specific case because it is population and trait dependent (Saatchi *et al.*, 2012).

The Spanish autochthonous cattle breeds have a *Bos taurus* ancestral origin and it is estimated that they have a common origin (Beja-Pereira *et al.*, 2003), with estimated $F_{ST}$ statistics between them ranging between 0.009 and 0.068 (Cañas-Álvarez *et al.*, 2015) and with a quite important persistency of haplotype phase between them (Cañas-Álvarez *et al.*, 2016). These characteristics jointly with the small size of these populations and the limited economic resources available for genotyping, suggest an appealing scenario to explore the implementation of genomic selection in a meta-population context. Thus, the objective of this study is to evaluate the efficiency of the potential application of multi-breed genomic selection in the Spanish beef cattle populations through a simulation study using a genomic data set to mimic the populations' LD structure.

## Material and methods

The genomic data used for simulation purposes consisted of genotypes from 171 triplets (sire/dam/offspring) obtained with the Illumina *BovineHD Beadchip* from seven breeds: *Asturiana de los Valles* (AV), $n = 25$, *Avileña - Negra Ibérica* (ANI), $n = 24$, *Bruna dels Pirineus* (BP), $n = 25$, *Morucha* (Mo), $n = 25$, *Pirenaica* (Pi), $n = 24$, *Retinta* (Re), $n = 24$ and *Rubia Gallega* (RG), $n = 24$. The triplets were sampled under the criteria of minimizing the genealogical relationship between them in order to capture as much of the variability as possible in each population.

The SNP filtering process included the following requirements: (1) Mendelian error bellow 0.05, (2) SNP and individual call rates higher than 95% and (3) minor allele frequency higher than 0.01, and it was performed using the *Plink* software (Purcell *et al.*, 2007). Further, SNPs that were located on the autosomal chromosomes were kept and those found in repetitive positions were excluded. At the end of the process there were 629 251 SNP markers covering 2 510 350 kb of the autosomal chromosomes with a mean density of one marker per 3.99 kb. The reconstruction of the parental haplotypes was conducted with the software *Beagle* (Browning and Browning, 2009) using the 'TRIO' option.

### *Simulation*

As mentioned, the simulation procedure was set up to reproduce the existing linkage disequilibrium structure of each breed. Thus, we started with the 100 available paternal haplotypes (50 individuals) for each breed comprising 629 251 SNP markers. Then, we expanded the populations to 500 individuals in the first generation. For each population, the 629 251 SNP markers of the individuals of the first generation were simulated by gene-dropping and assuming a map distance of 1 cM every Mb. Their parents were selected randomly from the previous generation (50 individuals) ignoring their sex. Following the same procedure, six more generations of 1000 individuals (100 sires and 900 dams) were simulated, selecting the parents randomly but considering their sex this time. The first three generations were used to establish the training populations for each breed, and the last three were used for validation. A summary of this simulation procedure is presented in Figure 1.

Further, in order to simulate the causative mutations of a trait, 3% of the SNP markers of each chromosome were randomly selected as QTL and they were attributed an additive effect sampled from a Gaussian distribution with zero mean and a standard deviation of one. Later on, for every individual, true breeding values (TBVs) were calculated as the sum of the effects of their genotype for the QTL polymorphisms. In addition, beside the polygenic model, five other scenarios of genetic architecture of the quantitative traits were simulated to create a sensitivity analysis:

- LMAF: Only markers with extreme frequencies (minor allele frequency $\leqslant 0.05$) were chosen to be QTLs.
- Ex: The effects of the QTLs were drawn from an exponential distribution instead of a Gaussian distribution.
- 10G(20%): 10 randomly selected QTLs which explained 20% of the total genetic variance were added to the polygenic model.
- 4G(50%): Four randomly selected QTLs were added to the polygenic model explaining 50% (5%, 10%, 15% and 20% for each QTL) of the total genetic variance.

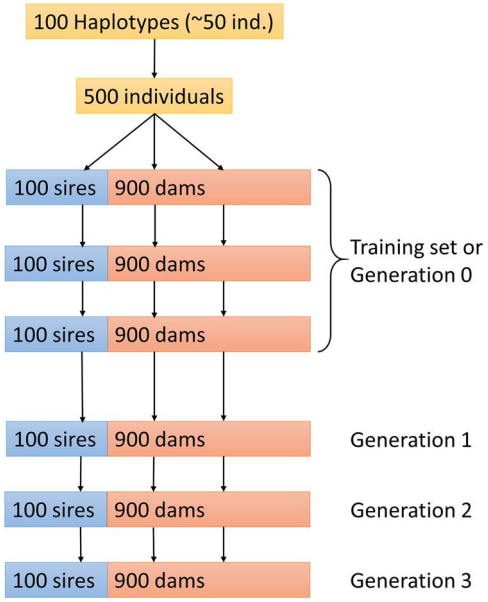


**Figure 1** Structure of the simulation strategy for the generation of pseudo-populations for each population.

- 4MG: Four QTLs were randomly selected to explain the 100% of the genetic variance with effects drawn from a normal distribution.

For each scenario, phenotypes were simulated for all individuals in the training population summing to their TBV, a trait mean (= 1000) and a residual drawn from a Gaussian distribution with appropriate variance to generate two traits with heritability 0.4 and 0.1, respectively.

*Genomic evaluation*

The genomic evaluation was performed by means of a Bayesian procedure with the *solveSNP* software (Legarra and Misztal, 2008) and under the following model.

$$y_i = \mu + \sum_{j=1}^{n} x_{ij} a_j + e_i$$

where $y_i$ is the phenotype of the $i$th individual, $\mu$ the trait mean, $n$ the number of SNPs, $x_{ij}$ the genotype of the $i$th individual for the $j$th marker coded as 0, 1 and 2, $a_j$ the substitution effect for the $j$th marker and $e_i$ the residual effect of the $i$th individual. Further, the prior distribution for the marker effects was the following multivariate Gaussian distribution:

$$\boldsymbol{a} \sim MVN\left(0, \boldsymbol{I}\sigma_a^2\right)$$

where $\sigma_a^2$ is the marker variance whose prior distribution was assumed to be uniform within appropriate bounds.

The SNP markers selected as causal mutations were excluded from the marker panel during the genomic evaluation. Later on, genomic estimated breeding values (GEBV) were calculated as:

$$GEBV_i = \sum_{j=1}^{n} x_{ij} \widehat{a}_j$$

Several scenarios of genomic evaluation were considered depending on the size and composition of the reference population.

- Pure-bred (PG): The reference population comprised 3000 individuals of one of the populations simulated. All seven populations were used as reference populations separately.
- Admixed_2 (×2): The reference populations comprised 3000 (1500 + 1500) randomly selected individuals from two purebred populations. All possible combinations were used as reference populations.
- Admixed_7 (×7): One reference population comprised 3003 individuals with 429 randomly selected individuals from each of the seven populations.

In addition, reduced purebred populations of 1500 and 429 individuals were used with the aim of comparing them to the admixed scenarios under equal genotyping effort within populations. As before, genotyped individuals were sampled randomly from the training populations.

*Validation*

In order to validate the predictions, we calculated the accuracy as the Pearson correlation between the simulated and the predicted breeding values in three additional generations (generations 1, 2 and 3) of 1000 non-phenotyped individuals for every population. Each case of simulation was replicated five times and we present the mean and standard error per generation.

## Results and discussion

*Single-breed evaluation*

In the first scenario, the effects of the SNP markers were estimated within each breed. Then, they were used to predict the GEBV within and across breeds. Figure 2 and Supplementary Material Table S1 show the results of the accuracies obtained for a trait with heritability 0.4 in all populations in generation 1 and for two subsequent generations. Within-breed accuracies at the first generation were the highest, ranging from 0.637 (RG) to 0.580 (BP). These results are similar to those reported by Saatchi *et al.* (2011) and Van Eenennaam *et al.* (2014) from empirical field studies with traits of comparable heritabilities. On the other hand, the across-breed accuracies were very low, with the highest value obtained when training in BP to predict over Pi (0.177) and the lowest when training in Pi to predict over Re (0.087). These results confirm the postulate of Harris *et al.* (2008) that indicated that training in one population and to predict in another is not effective. However, it is remarkable that all the average estimates of accuracy were positive which is coherent with the persistency of LD found between these populations and the genetics closeness between them (Cañas-Álvarez *et al.*, 2016).

When predicting the subsequent generations the within-breed accuracies resulted on average lower by 16.4% in generation 2 with values between 0.545 (RG) and 0.475 (AV) and 24.3% in generation 3 with values between 0.496 (Pi) and 0.420 (BP) with respect to generation 1. The decrease in accuracy across generations was expected and confirms the relevance of the relationship between the testing and training populations in the accuracy of genomic selection (Clark *et al.*, 2012). Nevertheless, the across-breed accuracies remained similar to generation 1, because the relationship between testing and training populations was not modified in these cases.

The results of accuracy obtained when evaluating for a trait with heritability 0.1 were similar but with lower magnitude of accuracies (Figure 3 and Supplementary Material Table S2). Thus, values ranged between 0.440 (Mo) and 0.380 (Re) for within-breed predictions and between 0.129 (AV over Mo) and 0.030 (BP over Re) for across-breed predictions in generation 1. The loss of predictive ability over subsequent generations resulted to be higher than that of the previous case. The within-breed accuracies were 18.1% lower in generation 2 (0.377 (Mo)–0.292 (BP)) and 30.0% in generation 3 (0.312 (Mo)–0.244 (BP)) than generation 1

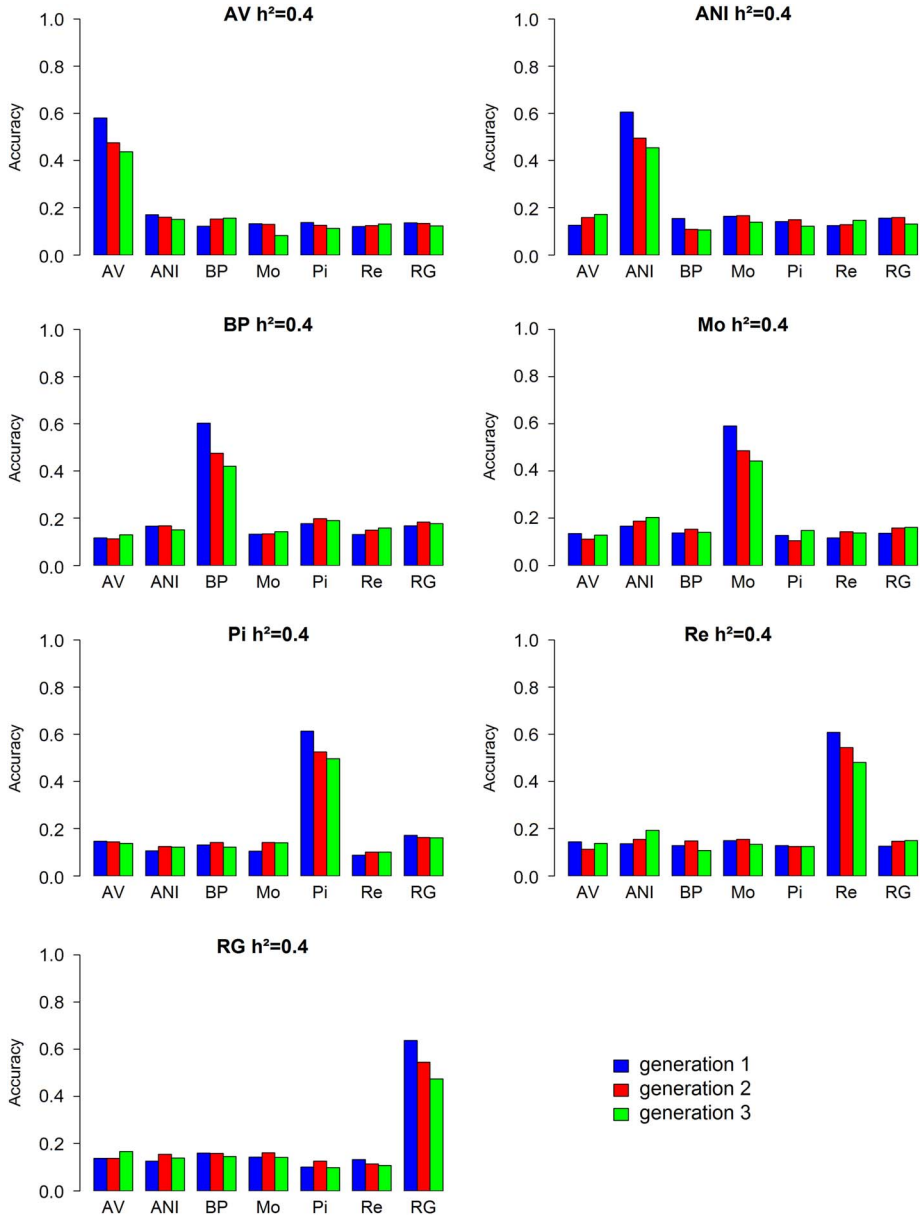


**Figure 2** Accuracy from single-breed genomic evaluation ($h^2 = 0.4$). AV = Asturiana de los Valles; ANI = Avileña-Negra Iberica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega; generations 1, 2, 3 = distance in generations between the training and validation sets.

(Figure 3). In traits with low heritabilities, the family information has a major impact on breeding value estimation (Falconer and Mackay, 1996). Therefore, as the relationship between testing and training populations weakened from one generation to the other, the loss of accuracy became more evident.

*Evaluation in admixed ×2*

The training sets used in this second scenario were set up by mixing data from two purebred populations with equal proportion of each. All possible combinations were considered which resulted in 21 different admixed populations. The results of the predictive ability of these populations over the purebred populations for generation 1 and for a trait with heritability of 0.4 are presented in Supplementary Material Table S3. When the purebred validation population was included in the admixed training set the accuracies ranged from 0.545 (AV–RG over RG) to 0.475 (AV–ANI over AV). However, when the purebred validation population was not included in the training set the accuracies resulted similar to the previous scenario of across-breed evaluation and ranging between 0.186 (AV–BP over RG) and 0.103 (Re–RG over Pi).

As an example, Figure 4 shows the results obtained from training in the AV–ANI population for all generations. As expected, the predictive ability over the subsequent generations was lower than that of the training population for AV and ANI. Though, the loss in predictive ability resulted slightly higher than that of the previous case. On average (Figure 5), the accuracies resulted 15.6% and 26.2% lower, compared with generation 1, for generations 2 and 3,

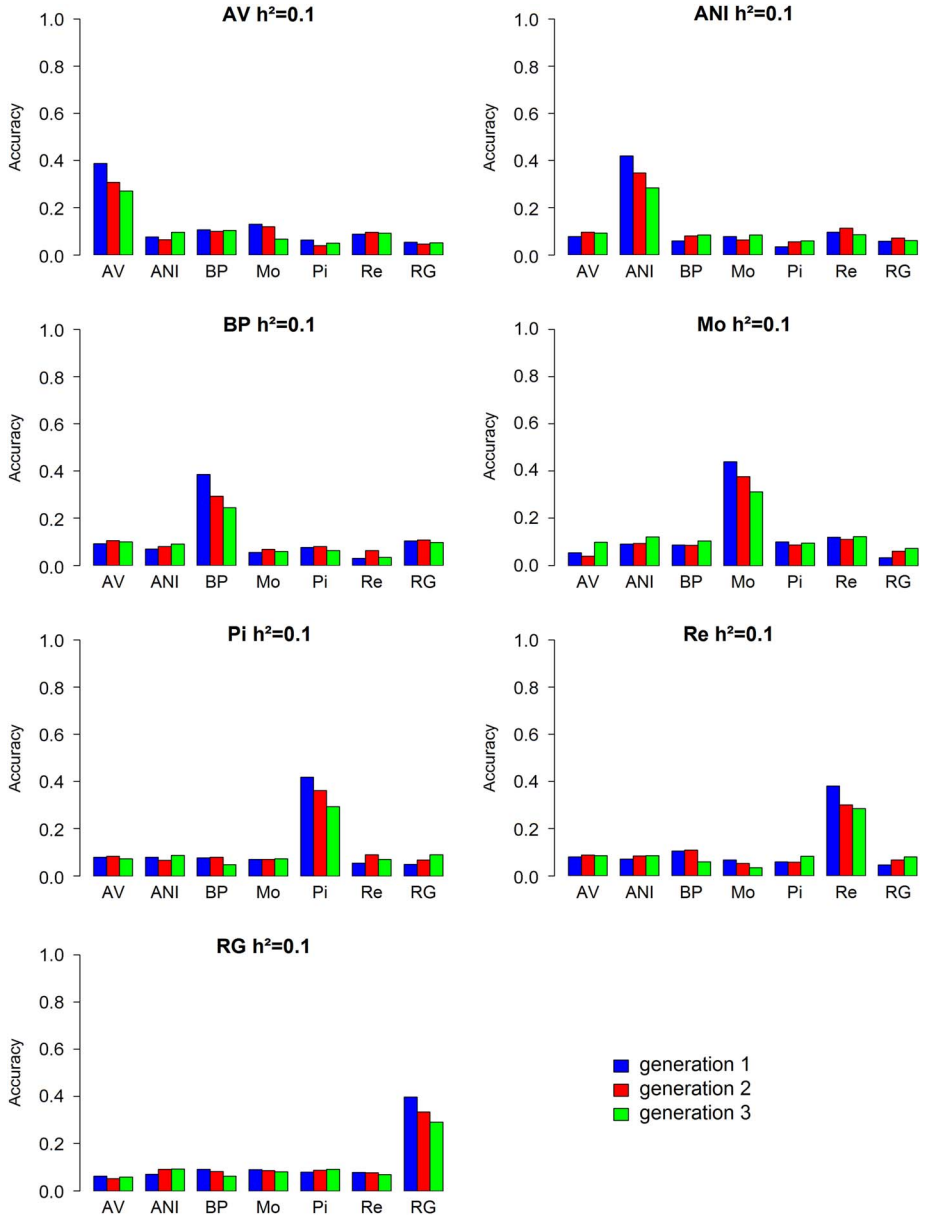


**Figure 3** Accuracy from single-breed genomic evaluation ($h^2 = 0.1$). AV = Asturiana de los Valles; ANI = Avileña-Negra Iberica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega; generations 1, 2, 3 = distance in generations between the training and validation sets.

respectively. The reason of this higher decrease of accuracy may be the limited number of individuals with a direct relationship with the testing population in the training set (1500 *v.* 3000).

Moreover, the results with heritability 0.1 were similar (Supplementary Material Table S4), although the overall accuracies resulted lower than previously. Accuracies for the populations included in the training set ranged between 0.338 (BP–Pi over Pi) and 0.290 (Pi–Re over Re). On the other hand, the results for the populations not included in the admixture were also lower, from 0.122 (AV–BP over Mo) to 0.029 (MO–Re over RG). The loss of accuracy in the subsequent generations was 17.7% and 30.2% for generations 2 and 3, respectively (Figure 5).

*Evaluation in admixed ×7*

Finally, the last training set used for genomic evaluation was constructed by combining data of 429 randomly selected individuals from each purebred population (total 3003 individuals). Figure 6 shows the results obtained for all populations and all generations. The accuracies resulted between 0.363 (BP) and 0.330 (AV) for the trait with $h^2 = 0.4$ and between 0.233 (Mo) and 0.159 (RG) for the trait with $h^2 = 0.1$, while the loss of accuracy with the generations was 16.1% and 22.2% for the first trait ($h^2 = 0.4$) and 15.8% and 29.7% for the second trait ($h^2 = 0.1$). The accuracies were lower than those in the previous scenarios and, as before, the loss of accuracy when training and testing populations were more distant was greater with lower heritability.

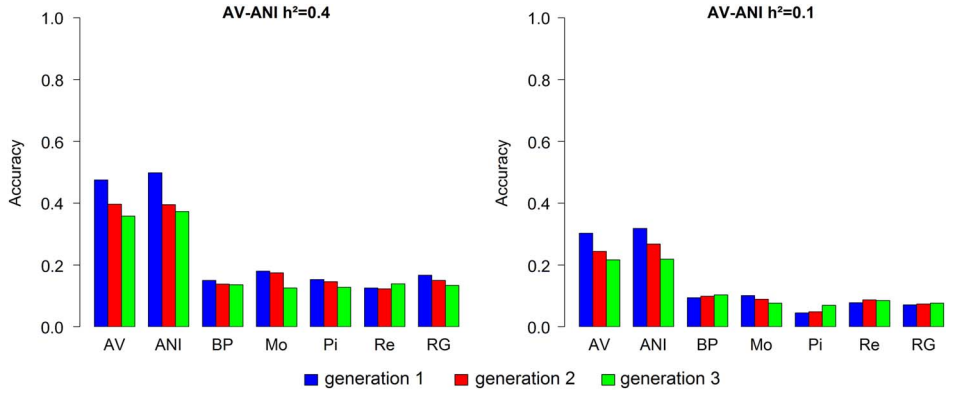


**Figure 4** Accuracy obtained from genomic evaluation in an admixed ×2 population (AV-ANI). AV = Asturiana de los Valles; ANI = Avileña-Negra Iberica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega; generations 1, 2, 3 = distance in generations between the training and validation sets.

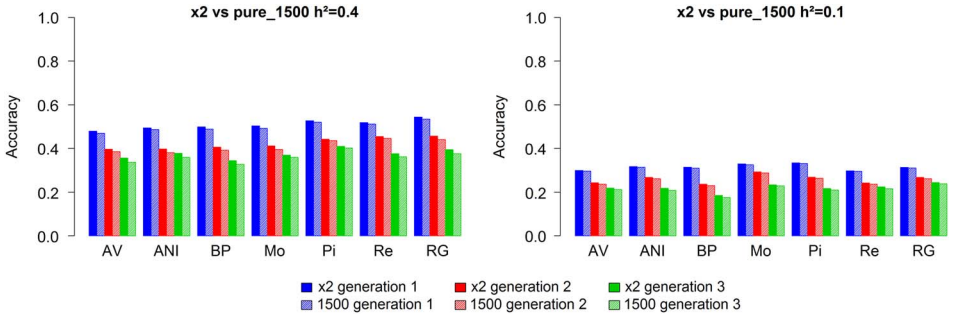


**Figure 5** Average accuracy obtained from genomic evaluation in the admixed ×2 population and comparison with the results of purebred genomic evaluation with 1500 individuals per population. AV = Asturiana de los Valles; ANI = Avileña-Negra Iberica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega; generations 0, 1, 2, 3 = distance in generations between the training and validation sets; ×2 = admixed training set from all two purebred populations (1500 + 1500 individuals); pure_1500 = purebred training set with 1500 individuals.

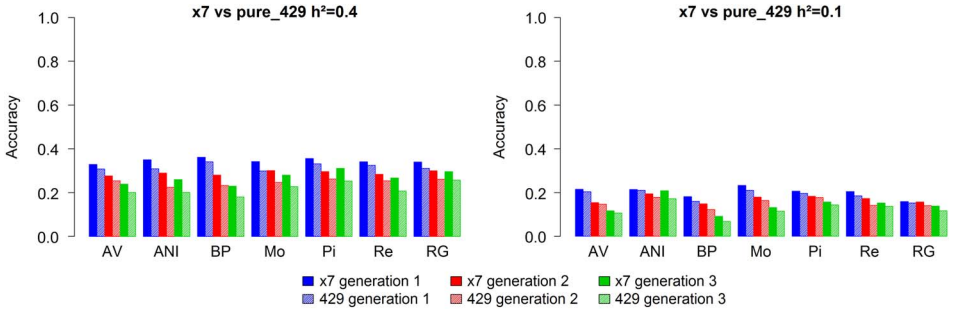


**Figure 6** Accuracy obtained from genomic evaluation in the admixed ×7 population and comparison with the results of purebred genomic evaluation with 429 individuals per population. AV = Asturiana de los Valles; ANI = Avileña-Negra Iberica; BP = Bruna dels Pirineus; Mo = Morucha; Pi = Pirenaica; Re = Retinta; RG = Rubia Gallega; generations 0, 1, 2, 3 = distance in generations between the training and validation sets; ×7 = admixed training set from all seven purebred populations (7 × 429 individuals); pure_429 = purebred training set with 429 individuals.

### *Admixed* v. *reduced purebred*

For each scenario, we also performed genomic evaluations using a subset of randomly selected individuals (1500 and 429) in order to compare them with the admixed populations and evaluate the effect of adding individuals from other populations to increase the size of the training dataset. Figure 5 (1500 individuals per population) and Figure 6 (429 individuals) present also the results of this comparison. They showed that adding information from another population to a small-sized training set was beneficial in all cases. The admixed ×2 populations performed slightly better than the reduced purebred populations with 1500 individuals with 1.7%, 3.0% and 4.2% higher accuracies for generations 1, 2, and 3, respectively, for a trait with $h^2 = 0.4$ and 1.1%, 2.1% and 3.4% for a trait with $h^2 = 0.1$. This superiority of the admixed population was more evident between the admixed ×7 and the reduced purebred populations of 429 individuals. Here, the gain in accuracy with the number of generations was 8.8%, 16.7% and 23.2% for the first trait ($h^2 = 0.4$) and 7.3%, 11.2% and 15.8% for the second trait ($h^2 = 0.1$). The most probable cause of this phenomenon is that as the relatedness between the training set and the validation set weakens the

predictions are based more on the short range LD between the markers and the QTLs than on the pure family relationship between individuals. Thus, the admixed populations perform better because of the higher number of data and the fact that mixing data breaks down the long distance LD created by relatedness and leaves the effects of the short range LD that persists through generations (Hill and Robertson, 1968). Nevertheless, these results indicate that the information provided by genetically related populations can be somewhat useful when the genotyping ability of a single population is restricted due to economic or size limitations and it seems to be more helpful when trying to predict individuals with weak genetic relationship with the training population.

*Genetic architecture of the trait*

Finally, we also compared the consequences of alternative genetic architecture of the traits. Thus, along with the polygenic traits (PG) simulated above, five more cases of genetic architecture were simulated as described earlier. In Supplementary Material Figure S1, we present the average within-breed accuracies for both traits ($h^2 = 0.4$ and $h^2 = 0.1$), obtained from training in purebred populations for all the populations simulated and for all generations. The values obtained were similar in all cases which is somewhat surprising for cases where QTLs explain a large amount of the genetic variance. As causal mutations are removed from the prediction process, the Bayesian procedure here used might not be the method of choice for these type of traits as it is unable to accommodate the differential effect of QTLs. To illustrate this phenomenon, Supplementary Material Figure S2 reflects the relationship between the LD with the causal polymorphism and the estimated effects for the closest 1000 markers in the validation and testing populations. The total effect of the QTL (20.16) was split in a large number of marker effects with estimates between −0.016 and 0.028. Some of them present a large LD with the causal polymorphism, but some others with a moderate LD are also associated with relevant SNP effects. Moreover, the relationship between LD and the estimated marker effects is reduced with the generations of separation between training and testing populations ($r^2 = 0.851$ in the testing population *v.* $r^2 = 0.796$ in the third generation of validation). Thus, if the existing LD between markers and causal mutations becomes smaller just by chance it would mean that the ability of markers to capture the effect of the QTLs would be reduced and therefore the effect on accuracy may become evident. Small differences can be observed only for the case of traits that are controlled by rare variants (LMAF) with MAF lower than 0.05 where the loss of accuracy was slightly greater with the number of generation.

Similarly, the results from the admixed ×2 and admixed ×7 training sets showed little differences among cases (Supplementary Material Figures S3 and S4, respectively). As before, only the LMAF case gives slightly lower accuracies. This phenomenon is coherent with the results obtained by Wientjes *et al.* (2015) that indicated that when the QTLs controlling the genetic variability of the traits have low frequencies the prediction ability of genomic selection is lower. However, although this has been suggested as the cause of the missing heritability (Gibson, 2012), the evidence for the percentage of genetic variation that rare variants produce is low and some authors have shown that these rare variants explain only a small percentage of the missing heritability of complex traits in human (Gusev *et al.*, 2014) or cattle (Gonzalez-Recio *et al.*, 2015).

Moreover, the reduced sized purebred populations performed similarly as in all previous cases. The only exception was that of the LMAF case where the reduced purebred training sets yielded higher accuracies than those of the admixed training sets with the number of generations. In the LMAF case, the markers selected to simulate the causal mutations where selected under the condition of having extreme frequencies (MAF $\leqslant 0.05$). As a consequence, the LD between the neutral markers and the QTLs is lower even at close distances and therefore, the reduced purebred training sets perform better than the admixed training sets because there is a larger proportion of family LD than sort range historical LD, even though the family relationship is decaying with the number of generations.

In conclusion, the results of this study indicate that the use of a meta-population may provide an increase of accuracy in scenarios with a reduced size of reference populations within breed. Further, the advantage of admixed populations seem to be greater when predicting individuals more distant from the training set because the across breed and multi-breed genomic predictions are based more on the LD between markers and causal polymorphism than on family relationships. This advantage should be more evident as the density of the genetic map growths. Nevertheless, some recent studies (Iheshiulor *et al.*, 2016; Van den Berg *et al.*, 2016) have probed that the use of full sequence data only provide marginal increases of the ability of prediction between distant breeds. Moreover, we assumed in this study that the genetic effect of QTL is constant across breeds, a typical assumption that may be wrong. The presence of epistatic or genotype × environmental interactions will also reduce the ability of prediction across breeds. Finally, it should also be noted that in this study all individuals in the reference populations were phenotyped and genotyped. The presence of missing phenotypes or genotypes in some individuals would probably reduce the ability of prediction.

## Acknowledgements

The authors want to thank the AGL 2010-15903 grant from the Spanish government; the collaboration of Breed societies in collecting samples and the support of FEAGAS is also acknowledged. J. J. Cañas-Álvarez acknowledges the COL-CIENCIAS support by the Francisco José de Caldas fellowship 497/2009 and A. González-Rodríguez acknowledges the financial support by the BES-2011-045434 fellowship.

## Supplementary material

To view supplementary material for this article, please visit https://doi.org/10.1017/10.1017/S175173111700283X

## References


Beja-Pereira A, Alexandrino P, Bessa I, Carretero Y, Dunner S, Ferrand N, Jordana J, Laloe D, Moazami-Goudarzi K, Sanchez A and Cañon J 2003. Genetic characterization of Southwestern European Bovine breeds: a historical and biogeographical reassessment with a set of 16 microsatellites. Journal of Heredity 94, 243–250.

Berry DP, Garcia JF and Garrick DJ 2016. Development and implementation of genomic predictions in beef cattle. Animal Frontiers 6, 32.

Bolormaa S, Pryce JE, Kemper K, Savin K, Hayes BJ, Barendse W, Zhang Y, Reich CM, Mason BA, Bunch RJ, Harrison BE, Reverter A., Herd RM, Tier B, Graser HU and Goddard ME 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. Journal of Animal Science 91, 3088–3104.

Browning BL and Browning SR 2009. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. American Journal of Human Genetics 84, 210–223.

Cañas-Álvarez JJ, González-Rodríguez A, Munilla S, Varona L, Díaz C, Baro JA, Altarriba J, Molina A and Piedrafita J 2015. Genetic diversity and divergence among Spanish beef cattle breeds assessed by a bovine high-density SNP chip. Journal of Animal Science 93, 5164–5174.

Cañas-Álvarez JJ, Mouresan E, Varona L, Diaz C, Molina A, Baro JA, Altarriba J, Carabano M, Casellas J and Piedrafita J 2016. Linkage disequilibrium, persistence of phase and effective population size in Spanish local beef cattle breeds assessed through a SNP high density chip. Journal of Animal Science 94, 2779–2788.

Chen L, Schenkel F, Vinsky M, Crews DH and Li C 2013. Accuracy of predicting genomic breeding values for residual feed intake in Angus and Charolais beef cattle. Journal of Animal Science 91, 4669–4678.

Clark SA, Hickey JM, Daetwyler HD and van der Werf JH 2012. The Importance of information on relatives for the prediction of genomic breeding values and the implications for the makeup of reference data sets in livestock breeding schemes. Genetics Selection Evolution 44, 4.

De Roos APW, Hayes BJ and Goddard ME 2009. Reliability of genomic predictions across multiple populations. Genetics 183, 1545–1553.

Falconer DS and Mackay TFC 1996. Introduction to quantitative genetics, 4th edition. Addison Wesley Longman Limited, Harlow.

Gibson G 2012. Rare and common variants: twenty arguments. Nature Reviews Genetics 13, 135–145.

Gonzalez-Recio O, Daetwyler HD, MacLeod IM, Pryce JE, Bowman PJ, Hayes BJ and Goddard ME 2015. Rare variants in transcript and potential regulatory regions explain a small percentage of the missing heritability of complex traits in cattle. PLoS One 10, e0143945.

Gunderson KL, Steemers FJ, Lee G, Mendoza LG and Chee MS 2005. A genome-wide scalable SNP genotyping assay using microarray technology. Nature Genetics 37, 549–554.

Gusev A, Lee SH, Trynka G, Finucane H, Vilhjalmsson BJ, Xu H, Zang C, Ripke S, Bulik-Sullivan B, Stahl E, Kahler AK, Hultman CM, Purcell SM, McCarroll SA, Daly M, Pasaniuc B, Sullivan PF, Neale BM, Wray NR, Raychaudhuri S and Price AL 2014. Partitioning heritability of regulatory and cell-type-specific variants across 11 common diseases. American Journal of Human Genetics 95, 535–552.

Harris BL, Johnson DL and Spelman RJ 2008. Genomic selection in New Zealand and the implications for national genetic evaluation. Proceedings of 36th ICAR Biennial Session, Niagara Falls, USA. p. 325.

Hayes BJ, Bowman PJ, Chamberlain AJ and Goddard ME 2009. Invited review: genomic selection in dairy cattle: progress and challenges. Journal of Dairy Science 92, 433–443.

Heidaritabar M, Wolc A, Arango J, Zeng J, Settar P, Fulton JE, O'Sullivan NP, Bastiaansen JW, Fernando RL, Garrick DJ and Dekkers JC 2016. Impact of fitting dominance and additive effects on accuracy of genomic prediction of breeding values in layers. Journal of Animal Breeding and Genetics 133, 334–346.

Hill WG and Robertson A 1968. Linkage disequilibrium in finite populations. Theoretical Applied Genetics 38, 226–231.

Iheshiulor OOM, Woolliams JA, Yu X, Wellmann R and Meuwissen THE 2016. Within – and across-breed genomic prediction using whole-genome sequence and single nucleotide polymorphism panels. Genetics Selection Evolution 48, 15.

Kachman SD, Spanger ML, Bennett GL, Hanford KJ, Kuehn L.A, Snelling WM, Thallman RM, Saatchi M, Garrick DJ, Schnabel RD, Taylor JF and Pollak EJ 2013. Comparison of molecular breeding values based on within- and across-breed training in beef cattle. Genetics Selection Evolution 45, 30.

Karoui S, Carabaño MJ, Díaz C and Legarra A 2012. Joint genomic evaluation of French dairy cattle breeds using multiple-trait models. Genetics Selection Evolution 44, 10.

Kizilkaya K, Fernando RL and Garrick DJ 2010. Genomic prediction of simulated multibreed and purebred performance using observed fifty thousand single nucleotide polymorphism genotypes. Journal of Animal Science 88, 544–551.

Legarra A and Misztal I 2008. Technical note: computing strategies in genome-wide selection. Journal of Dairy Science 91, 360–366.

Legarra A, Robert-Granié C, Manfredi E and Elsen JM 2008. Performance of genomic selection in mice. Genetics 180, 611–618.

Lourenco DAL, Tsuruta S, Fragomeni BO, Masuda Y, Aguilar I, Legarra A, Bertrand JK, Amen TS, Wang L, Moser DW and Misztal I 2015. Genetic evaluation using single-step genomic best linear unbiased predictor in American Angus. Journal of Animal Science 93, 2653–2662.

Meuwissen THE, Hayes BJ and Goddard ME 2001. Prediction of total genetic value using genome-wide dense marker maps. Genetics 157, 1819–1829.

Purcell S, Neale B, Todd-Brown K, Thomas L, Ferreira MAR, Bender D, Maller J, Sklar P, de Bakker PIW, Daly MJ and Sham PC 2007. PLINK: a tool set for whole-genome association and population-based linkage analyses. The American Journal of Human Genetics 81, 559–575.

Reiner-Benaim A, Erza E and Weller JI 2017. Optimization of a genomic breeding program for a moderately sized dairy cattle population. Journal of Dairy Science 100, 1–3.

Saatchi M, McClure MC, McKay SD, Rolf MM, Kim J, Decker JE, Taxis TM, Chapple RH, Ramey HR, Northcutt SL, Bauck S, Woodward B, Dekkers JCM, Fernando RL, Schnabel RD, Garrick DJ and Taylor JF 2011. Accuracies of genomic breeding values in American Angus beef cattle using K-means clustering for cross-validation. Genetics Selection Evolution 43, 40.

Saatchi M, Schnabel RD, Rolf MM, Taylor JF and Garrick DJ 2012. Accuracy of direct genomic breeding values for nationality evaluated traits in US Limousin and Simmental beef cattle. Genetics Selection Evolution 44, 38.

Silva RMO, Fragomeni BO, Lourenco DAL, Magalhaes AFB, Irano N, Carvalheiro R, Canesin RC, Mercadante MEZ, Boligon AA, Baldi FS, Misztal I and Albuquerque LG 2016. Accuracies of genomic prediction of feed efficiency traits using different prediction methods in an experimental Nelore cattle population. Journal of Animal Science 94, 3613–3623.

Toosi A, Fernando RL and Dekkers JCM 2010. Genomic selection in admixed and crossbred populations. Journal of Animal Science 88, 32–46.

Tussell L, Gilbert H, Riquet J, Mercat MJ, Legarra A and Larzul C 2016. Pedigree and genomic evaluation of pigs using a terminal-cross model. Genetics Selection Evolution 48, 32.

Vallejo RL, Leeds TD, Gao G, Parsons JE, Martin KE, Evenhuis JP, Fragomen BO, Wiens GD and Palti Y 2017. Genomic selection models double the accuracy of predicted breeding values for bacterial cold water disease resistance compared to a traditional pedigree-based model in rainbow trout aquaculture. Genetics Selection Evolution 49, 17.

Van den Berg I, Boichard D and Lund MS 2016. Sequence variants selected from a multi-breed GWAS can improve the reliability of genomic predictions in dairy cattle. Genetics Selection Evolution 48, 83.

Van Eenennaam AL, Weigel KA, Young AE, Cleveland MA and Dekkers JCM 2014. Applied animal genomics: results from the field. Annual Review of Animal Biosciences 2, 105–139.

Weber KL, Thallman RM, Keele JW, Snelling W M, Bennett GL, Smith TPL, McDaneld TG, Allan MF, Van Eenennaam AL and Kuehn LA 2012. Accuracy of genomic breeding values in multibreed beef cattle populations derived from deregressed breeding values and phenotypes. Journal of Animal Science 90, 4177–4190.

Wientjes YC, Calus MP, Goddard ME and Hayes BJ 2015. Impact of QTL properties on the accuracy of multi-breed genomic prediction. Genetics Selection Evolution 47, 42.