

## Design of Reference Populations for Genomic Selection in Crossbreeding Programs.<sup>1</sup>

E.M. van Grevenhof<sup>1</sup> and J.H.J. van der Werf<sup>2</sup>

<sup>1</sup>Animal Breeding and Genomics Centre, Wageningen University, The Netherlands,

<sup>2</sup>School of Environmental and Rural Science, University of New England, Armidale, Australia.

**ABSTRACT:** In crossbreeding (CB) programs, genomic selection offers the opportunity to make efficient use of information on crossbreds in the selection of purebred candidates (PB). In a deterministic simulation study, we evaluated the use of various proportions of crossbred animals in a reference population for genomic selection of purebred animals used in a crossbreeding program. When the breeding objective is to improve crossbred performance, the optimal proportion of CB in the reference population is 100% unless the correlation between purebred and crossbred performance is high. However, the value of replacing purebred animals with crossbred animals in the reference population is generally small unless the correlation between purebred and crossbred performance is relatively low ( $< 0.7$ ) or when purebred performance is not considered in the breeding objective.

**Keywords:** Genomic selection; Reference population; Purebred; Crossbred

### Introduction

Traits expressed in purebreds are genetically not the same as traits expressed in crossbreds (Wei et al., 1995). Purebred and crossbred performance can be considered as two genetically correlated traits, where the correlation ( $r_{pc}$ ) is affected by the extent of non-additive genetic effects (particularly dominance) and the genetic distance between lines or breeds crossed. The correlation often also reflects genotype by environment interaction due to nucleus purebreds (PB) and commercial crossbreds (CB) performing in different environments. Wei and van der Werf (1994) proposed to consider purebred and crossbred performance as two correlated traits and using a multi trait model for genetic evaluation of purebred and crossbred performance. The breeding objective usually focuses on the performance of CB animals. However, it is hard to use crossbred performance records as they can be difficult to obtain and selection for CB performance is either on half sibs (HS), which does not exploit variation within the HS family, or on progeny information, which would lengthen generation interval. Bijma and van Arendonk (1998) showed that extensive use of sib information on crossbreds can lead to increased rates of inbreeding. Genomic information allows information on CB to become available at an early age and to utilize within family variation. Dekkers (2007b) proposed to use genetic marker information which is calibrated based on the performance of commercial crossbred animals. He found a significant increase in rates of gain compared to using only purebred phenotypic information, or combined purebred and crossbred information, whereas the rate of inbreeding was decreased.

Genomic selection uses marker genotypes and phenotypes in a reference population to predict breeding values of selection candidates that have been genotyped (Meuwissen et al., 2001). The effectiveness of genomic selection will depend on the size and composition of the reference population used for genomic predictions (Daetwyler et al., 2008, Goddard, 2009). In crossbreeding programs, the breeding objective often includes crossbred performance, or in some cases, both PB and CB performance. Therefore, it seems recommendable to include performance and genotypic data on crossbred individuals for genomic selection in crossbreeding programs. However, collection of CB information might be difficult, expensive and time consuming. Moreover, large amounts of phenotypic as well as genotypic information of purebreds will usually be already available. The question then becomes, what is the additional benefit of adding CB to a reference population and how much accuracy is lost if we more pragmatically use a reference population that combines purebred and crossbred information? This study will assess the benefit of including CB information in the reference population of a breeding program using genomic selection. The efficiency of investing in CB information to enable genomic selection of purebred animals for CB performance, and the optimal balance between including PB and CB animals in a reference population is explored.

### Material and Methods

Selection index methodology provides a good framework for predicting the accuracy of estimated breeding values for various breeding program scenarios, including the use of genomic information (Dekkers, 2007a). Deterministic simulation was used to predict accuracy of estimated breeding value of selection candidates in a two-way crossing system. Purebred animals are selected for an index that includes varying amounts of records and genotypes on crossbred and purebred individuals. In scenarios with genomic selection, the make-up of the reference population was varied in terms of size, and proportion of crossbred individuals.

**Breeding program.** We assume a pig breeding nucleus with each of two purebred lines containing 500 breeding females and 25 females mated to one male. The weaning rate is 11.5. From each full sib family, 2 males and 2 females will be measured to replace nucleus parents. Nucleus replacements are selected on a single trait index measured at fixed time of selection. At time of selection, phenotypic information is available on the own performance, sire and dam, 4 full sib records and 40 half sib records of purebred family members. Purebred animals can

also be genotyped and reliability of their genomically estimated breeding value is dependent on the size and composition of the reference population. Purebred animals are mated to produce crossbreds and the breeding goal is to improve the performance of crossbreds, within some breeding schemes also purebred performance having some economic value. Purebreds have crossbred relatives phenotypically measured at the time of their selection (10 crossbred half sibs, no crossbred progeny). We will consider the PB performance and the CB performance as two genetically correlated traits, with correlation  $r_{PC}$ .

**Selection index.** Following Dekkers (2007a, 2007b), multi trait selection indices were derived to predict  $g$ , the breeding value of the purebred animal for performance of crossbred progeny, i.e. crossbreeding value. We combined phenotypic information with genomic predictions of breeding value, i.e. the molecular breeding value ( $mbv$ ), from both PB and CB information sources. The selection index was optimized for the breeding objective with varying emphasis on PB and CB performance. The accuracy of index was used to compare the performance of different scenarios.

**Accuracy of genomic breeding value.** The accuracy  $r_{g\hat{g}}$  of the estimated breeding values based on genomic information ( $mbv$ ) was derived from the size and composition of total reference population. We considered two reference populations, one for PB and one for CB of size  $(1-p_{CB}) \cdot n_p$  and  $p_{CB} \cdot n_p$ , respectively, where  $p_{CB}$  is the proportion of crossbred animals in the reference population and  $n_p$  is the size of the combined reference population. We used the formula of Daetwyler et al. (2008) to predict  $r_{g\hat{g}}$  of both  $mbv_{PB}$  and  $mbv_{CB}$ . For the crossbred reference population, each individual is assumed to be half as informative, as genomic information can be based on only one rather than two haplotypes, i.e. only half as many relevant chromosome segments. We assumed an effective size of the breeding population of  $N_e = 100$ .

Basic parameters are described in Table 1. We varied the size of the reference population ( $n_p$ ), the proportional emphasis on PB in the breeding objective (PB\_EV), and the correlation between PB and CB performance ( $r_{PC}$ ). In all comparisons we plotted the index accuracy versus  $p_{CB}$ .

**Table 1. Base parameters used for simulation.**

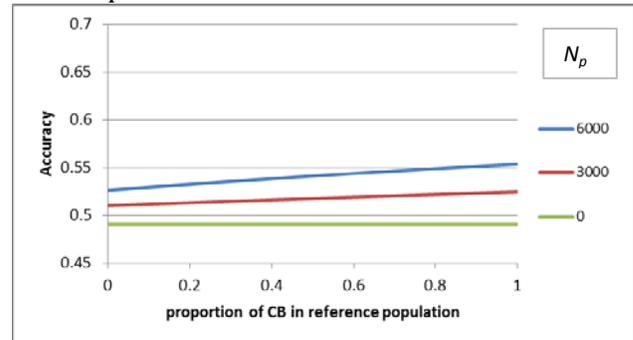
	PB / CB
Heritability, $h^2$	0.25
Phenotypic standard deviation, SD	1
Common environment among full sibs, $c^2$	0.15
Economic value, EV	0 / 1
Effective population size, $N_e$	100
Reference population size, $n_p$	4,000
Purebred-Crossbred correlation, $r_{PC}$	0.7

## Results

**Reference population structure.** The selection accuracy based on purebred phenotypic information alone was 0.45. Combining phenotypic information on PB and CB gave an accuracy of 49%, i.e. 10% higher than using

purebred information alone. When additionally using genomic selection and using a PB reference population of size 6000 the accuracy increased to 0.53. In the most desirable scenario of using genomic selection based on 6000 crossbred individuals, the accuracy was 0.55. In general, increasing the proportion CB animals in the reference population will increase the accuracy (Figure 1), but this increase is relatively small in the base scenario. When using purebred animals in the reference population, the accuracy is 7% higher with genomic selection compared with just phenotypic selection on PB/CB information. This accuracy increase is 13% if a CB reference population is used. For a large reference population ( $N_p = 6000$ ) the difference between using a PB reference population versus a CB reference population in accuracy is 5.2%. The increase is nearly linear, so a 50% CB/PB reference population gives about 2.8% more gain than a PB reference population.

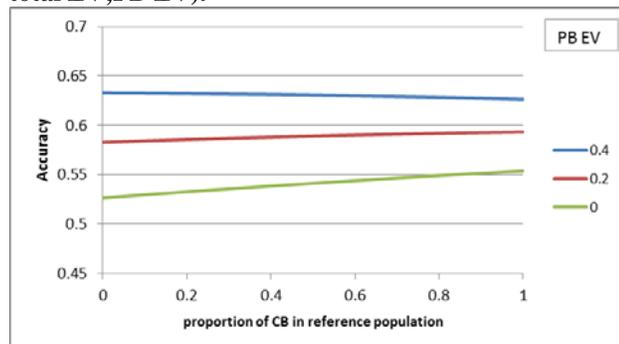
**Figure 1. Effect of CB proportion in reference population ( $N_p$ ) on selection accuracy of PB for varying size of  $N_p$**



**Economic value.** Figure 2 compares the accuracy for different breeding objectives where the relative emphasis on purebred performance (PB\_EV) varies from 0 to 0.4. When the PB performance has some economic value in the breeding objective (PB\_EV>0), then the value of having a CB animals in the reference population becomes smaller (Figure 2). While the value of a CB over a PB reference population is 5% if the PB\_EV=0, this decreases to 2% if PB\_EV=0.2 whereas there is even a small (1%) loss if the PB\_EV=0.4. The reason for the loss is that information on genomic prediction accuracy delivered per CB animal is lower as it only contains information from one haplotype of the purebred line. The accuracy of a genomic prediction from 6000 PB animals was predicted to be 0.46 whereas this was only 0.33 when predicting from 6000 CB animals.

**Correlation between PB and CB performance.** Figure 3 shows the effect of the  $r_{PC}$  value on the selection accuracy and how it relates to optimal  $p_{CB}$ . For a high  $r_{PC}$  value of 0.9 the effect of replacing PB by CB in the reference population is negative, again because of the lower information contained by the single haplotype of the CB animal. The difference between using a PB reference population versus a CB reference population is 5.2% for  $r_{PC}=0.7$  and it is 14.8% for  $r_{PC}=0.5$ . Again, the increase is nearly linear, so with a 50% PB/CB reference population, this additional accuracy is about halved.

**Figure 2. Effect of CB proportion in reference population on selection accuracy of PB for varying emphasis of PB in breeding objective (proportion of total EV;PB EV).**



**Effective population size.** With an  $N_e$  value of 500, the added value of genomic selection based on a PB reference population is only 1.7% and this increase is 2.9% if the reference population consists of only CB animals. The accuracy of a genomic prediction from 6000 PB animals was predicted to be 0.22, whereas this was only 0.16 when predicting from 6000 CB animals. Both predictions assume no direct pedigree relationships between reference population and selection candidates and probably underestimate the accuracy achieved in practice.

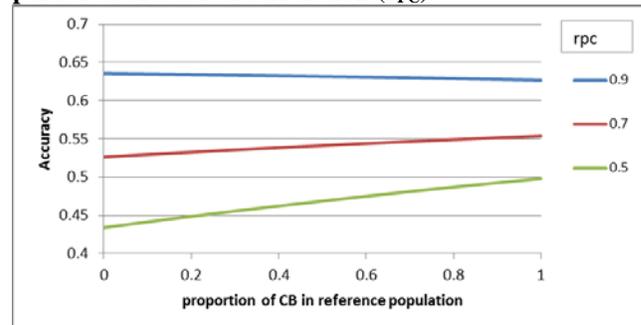
### Discussion

Genomic selection can be very valuable in crossbreeding programs as it allows efficient selection on crossbred performance. In general, a larger proportion of CB animals in the reference population will increase the selection accuracy when the breeding goal is focussed on CB performance. However, it might be difficult and expensive to collect phenotypes and genotypes on crossbred individuals, whereas most breeding programs have routine phenotyping measurement and genotyping of nucleus animals in the pure lines. We found that the effect of replacing purebred animals with crossbred animals in the reference population has some value. When the correlation between purebred and crossbred performance is low ( $<0.7$ ) and the breeding objective emphasis is mainly focused on improving CB, then higher values for  $p_{CB}$  lead to higher accuracy. In practice, breeding objectives often have some emphasis on PB performance, in which case the added value of additional CB animals in the reference becomes very small. The value of increasing the percentage of CB in the reference population depends on the  $r_{PC}$  as well as on  $PB\_EV$ . If  $r_{PC}$  values are not very low ( $\sim 0.8$ ) and  $PB\_EV$  is  $>0$ , which is a realistic scenario, then the benefit of using PB animals in the reference population is similar than that of CB animals.

In this study, we did not account for the Bulmer effect, which would have given a reduced value to the phenotypic information on CB half sib (van Grevenhof et al., 2012). Hence, the value of genotyping CB rather than just using their phenotypes will be higher in reality than presented in this paper. Also, the effect on inbreeding will

be favourable when applying GS in crossbreeding programs. These advantages have been pointed out by Dekkers (2007a). The main focus of this paper was to look at the effect of different proportions of PB and CB in the reference population, when applying genomic selection in CB programs, and these would not be much affected by the Bulmer effect.

**Figure 3. Effect of CB proportion in reference population on selection accuracy of PB for varying purebred-crossbred correlation ( $r_{PC}$ ).**



### Conclusions

Genomic selection for CB performance can significantly increase rates of genetic gain in crossbreeding programs. Having crossbred animals in the reference population gives more gain than having just a PB reference population. However, the value of replacing purebred animals with crossbred animals in the reference population is generally small unless the correlation between purebred and crossbred performance is low ( $<0.7$ ) or when purebred performance is not considered in the breeding objective.

### Acknowledgements

This work was supported by Science and Technology Foundation of the Netherlands Organization for Scientific Research (NWO-STW), with co-financers Institute for Pig Genetics (TOPIGS BV) and Product Board Animal Feed (PDV).

### Literature Cited

- Bijma, P. & Arendonk, J.A.M. (1998). *Anim Sci* 66: 529.
- Daetwyler, H.D., B. Villanueva, and Woolliams, J.A. (2008). *PLoS One* 3(10): e3395.
- Dekkers, J.C.M. (2007a). *J. Anim Sci* 85: 2104-2114.
- Dekkers, J.C.M. (2007b). *J. Anim Breed Genet* 124: 331-34.
- Goddard, M.E. (2009). *Genetica* 136: 245-257.
- Meuwissen, T., B. Hayes and Goddard, M. (2001). *Genetics* 157: 1819-1829.
- Van Grevenhof, E.M., Arendonk, J.A.M. and Bijma, P. (2012). *Gen Sel Evol.* 44: 26
- Wei, M. and van der Werf, J.H.J. (1994). *Anim Prod* 59: 401-413.
- Wei, M. and van der Werf, J.H.J. (1995). *J of Anim Sci* 73: 2220-2226.