

Diminishing marginal returns from genomic selection as more selection candidates are phenotyped

T. O. Okeno¹, M. Henryon^{2,3} and A. C. Sørensen¹

¹Center for Quantitative Genetics and Genomics, Department of Molecular Biology and Genetics, Aarhus University, Denmark, ²Danish Agriculture and Food Council, Pig Research Centre, Copenhagen, Denmark and ³School of Animal Biology, University of Western Australia, Crawley, Australia.

ABSTRACT: We used stochastic simulation to test hypotheses that, (i) phenotyping proportion of high ranking selection candidates based on estimated breeding values (EBV) before genotyping could realize as much genetic gains as phenotyping all candidates, and (ii) there is diminishing return to selection as more candidates are phenotyped in genomic breeding programs. Three phenotyping criteria, namely, random (RS), EBV and true breeding value (TBV) were investigated under two schemes (across-population and within-litter) using traditional-BLUP and genomic-BLUP models. The EBV ranked above RS and realized maximum achievable gains in the breeding program at 80% phenotyping. There was diminishing return to selection as more candidates were phenotyped; indicating that information content is not so large in the last candidate compared to the first candidate phenotyped. These findings demonstrate the need to rank selection candidates and phenotype only top 80% in genomic selection program as no bias was detected.

Keywords: Phenotyping; pigs

Introduction

Several studies have demonstrated the benefit of genomic selection in pig breeding in terms of increased rates of gain (Lillehammer et al. (2011); Wellmann et al. (2013)). Generally, implementation of genomic selection in breeding programs requires genotyping of selection candidates as well as establishment of reference population(s) of phenotyped and genotyped animals. In practice, and especially in fecund species like pigs, there are many candidates in a breeding program and decisions need to be made on which and how many candidates to phenotype and genotype.

In a simulation study, Henryon et al. (2012) showed that genotyping 5 to 20% of the candidates would realize most of the benefits of genomic selection with 100% genotyping. In that study, the assumption was that all the candidates had phenotypic information, which is consistent with the current practice for traits like early growth rate. Phenotypic information is important because it increases the accuracy of predicted breeding values of phenotyped candidates and such candidates can be added to the reference population, thereby increasing the accuracy of all selection candidates. Although, in practice, all candidates are always phenotyped in breeding programs, we envision that we could realize most benefits of genomic selection by phenotyping only proportion the candidates just as in genotyping (Henryon et al. (2012)). We reason that, if parental generation information is available, then such information could be utilized to estimate parents' averages of selection candidates. Then, high ranking candidates are

selected for phenotyping before genotyping. Such candidates are expected to generate higher breeding values because they are more likely to have large true breeding values compared to low ranking ones and are more likely to be selected for breeding. All candidates do not have similar genetic merits and, therefore, the last candidate phenotyped, i.e. the lowest ranking, may not be as valuable as the first candidate phenotyped for two reasons. First, the information in the phenotype is unlikely to increase the selection criteria of the last candidate above the selection threshold, and therefore, change selection decisions, especially in species with high selection intensity. Second, the phenotype of the highest ranking animal is likely more valuable in the future reference population. This could result in diminishing return to selection as more candidates are phenotyped similarly to the case of genotyping as demonstrated by Henryon et al. (2012). Based on these arguments we hypothesized that: (i) phenotyping only a proportion of high ranking selection candidates before genotyping is almost as good as phenotyping all the available candidates and (ii) there is diminishing marginal return from genomic selection when more candidates are phenotyped. We used stochastic simulation of breeding scheme that resembled that used in pigs to investigate these hypotheses.

Materials and Methods

Procedure: We used stochastic simulation to predict genetic gains in a genomic selection program, where 20-100% of the selection candidates were phenotyped before all of the selection candidates were genotyped.

Phenotyping criteria: Three criteria were used to choose which candidates to phenotype. They included:

- *True breeding value (TBV):* The assumption in this criterion was that, we have perfect knowledge on the genetic merits of the selection candidates and, therefore, all the available candidates were ranked based on their true breeding values and the highest ranking candidates phenotyped.
- *Estimated breeding value (EBV):* This criterion assumed that, the pedigree and genomic information were available from the parental generation. This information was used to estimate the parent averages of the selection candidates. The candidates were ranked based on estimated breeding values and the highest ranking candidates phenotyped.
- *Random selection (RS):* Selection candidates were randomly sampled from all the available selection candidates for phenotyping.

Breeding scheme: A breeding program with 12 discrete generations and fixed numbers of boars and sows per generation was considered. Ten boars were randomly mated to 100 sows with each sow farrowing five piglets per litter, resulting to 500 piglets per generation. All the selection candidates were assumed to be genotyped after they were phenotyped. Selection was for a single trait with heritability $h^2=0.4$ and genetic variance of 1. All the scenarios were the same in the first seven generations where traditional BLUP was used and all candidates phenotyped. From generation 8, the scenarios differed as the selection models traditional-BLUP or genomic-BLUP and phenotyping criteria (RS, EBV and TBV) were used. In each criterion, the phenotyping candidates were either selected across the population (AP) or within litters (WL). Six phenotyping strategies based on phenotyping criteria and selection unit were therefore modeled and evaluated for 5 generations. They were categorized as AP (RS-AP, EBV-AP and TBV-AP) and WL (RS-WL, EBV-WL and TBV-WL) strategies. Since selection based on RS-WL and EBV-WL was identical to RS-AP, they were ignored and therefore only the three AP strategies and one WL strategy (TBV-WL) were considered.

Simulated founder population: The genetic architecture of the founder population was generated to reflect the linkage disequilibrium observed in the Danish pigs. The founder population was simulated from 50 animals (25 boars and 25 sows). Random mating was applied in this population for 1000 generations. The genome was assumed to consist of 18 chromosomes of 167 cM each. Each chromosome contained 3.0×10^8 loci evenly distributed across the genome. The rate at which the original allele converted to a new allele was 4.0×10^{-6} and every 8th locus was a potential QTL. After 1000 generations, a total of 58,572 markers and 8,302 QTL were segregating. Effects of QTL alleles were sampled from a normal distribution with mean zero and afterwards scaled to give a genetic variance of 1. Haplotypes from the last founder generation were sampled to initiate the breeding schemes.

Prediction of breeding values: The breeding values were predicted using two evaluation methods. The first method used the traditional best linear unbiased prediction procedure (BLUP) to predict the estimated breeding values which were used as bench mark for comparison purposes. In this method, no genotyping was carried out. The BLUP model was:

$$y = Xb + Zu + e \quad (1)$$

where y is the vector of observations, X and Z are incidence matrices for the vectors of parameters b and u , u is the vector of breeding values of animals with variance $Var(u) = A\sigma_a^2$, A is the numerator relationship matrix, and σ_a^2 the additive genetic variance equal to 1. The e is the vector of random residual errors with $Var(e) = I\sigma_e^2$,

where I is the identity matrix and σ_e^2 is the error variance set to 1.5.

The second method used single-step genomic-BLUP (GBLUP) procedures as described by Gao et al. (2012) to predict the genomic breeding values. The GBLUP model was:

$$y = Xb + Zg + e \quad (2)$$

where y , b , g , and e were the vectors of trait phenotype, fixed effects, random effects and random residual errors, while X and Z are incidence matrices. The fixed effects included generational effects and overall mean. The g is the vector of genomic breeding values of animals with variance $Var(u) = H\sigma_a^2$, where H is a combination of A and a genomic relationship matrix calculated according to the method by VanRaden (2008) and σ_a^2 the additive genetic variance, equal to 1. The e is the vector of random residual errors with $Var(e) = I\sigma_e^2$ where I is the identity matrix and σ_e^2 is the error variance, set equal to 1.5.

Data Analysis: We used the stochastic simulation program ADAM (Pedersen et al. (2009)) to test the phenotyping strategies. Each strategy was replicated 100 times. The genetic gain for trait in the breeding objective was calculated as a regression of generation means in generations 8-12.

Results and Discussion

TBV-WL was found to be higher in genetic gains than AP strategies in both BLUP and GBLUP models (Figure 1). In GBLUP, for example, the genetic gains in TBV-WL were 1.1-4.2% higher than AP strategies with 20-60% phenotyping. This could be explained by higher accuracy (0.63) observed in WL compared to AP (0.48). The superiority of WL selection has been reported earlier (Lillehammer et al. (2013)). In that study, the accuracy of genomic selection was found to be higher with within-family selection than across population selection.

The genetic gains achieved in the AP strategies before 100% phenotyping differed significantly ($p < 0.05$), indicating that phenotyping strategy chosen has an impact on genomic selection program. Generally, EBV-AP was superior in genetic gains, while TBV-AP ranked lowest under GBLUP model. On the other hand, genetic gains realized in TBV-WL surpassed those obtained in AP strategies with 20-60% phenotyping. The low ranking of TBV-AP was unexpected as we reasoned that, having perfect knowledge on an individual's genetic merit could give us better idea when choosing individuals to phenotype. However, the findings demonstrate that, such information does not add much value in genomic selection, but could be important in traditional BLUP selection as TBV-AP and TBV-WL ranked above EBV-AP and RS-AP strategies when selection was based on traditional BLUP (Figure 1).

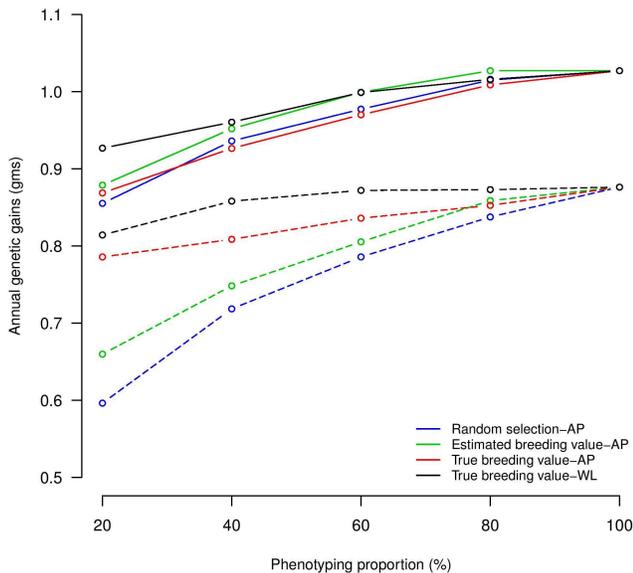


Figure 1. Annual genetic gains for the four phenotyping strategies under traditional BLUP (----) and GBLUP (—) models

The genetic gains realized by phenotyping 80% of the candidates was found to be significantly different ($p < 0.05$) from 100% phenotyping when RS-AP was employed, but had no difference with EBV-AP strategy under GBLUP. The fact that EBV-AP ranked above RS-AP (Figure 1) and realized maximum genetic gain achievable with only 80% phenotyping, demonstrate its worth over RS-AP. This means that if RS-AP is used then, the remaining 20% of the candidates must be phenotyped to achieve maximum genetic gain, while in EBV-AP, phenotyping those candidates will not add much benefit to the breeding program. These findings confirm our reasoning that, we can achieve most of the benefit from genomic selection by phenotyping just a proportion of available candidates, provided we can utilize pedigree information to rank them based on estimated breeding values and phenotype only high ranking ones. This would have a positive impact on cost reduction in a breeding program.

Our hypothesis that the marginal returns in genetic gain diminish as we increase the proportions of phenotyped selection candidates in genomic breeding programs was confirmed by the results. The genetic gains increased at diminishing rate in all four strategies, but mostly so for EBV-AP (Figure 1). For EBV-AP in GBLUP model, the genetic gains increased by 0.073, 0.046, 0.030, and 0.001 as phenotyping proportion increased from 20 to 40, 40 to 60, 60 to 80 and 80 to 100, respectively. These findings concur with those reported by Henryon et al. (2012), where they demonstrated that as more candidates were genotyped after phenotyping, the genetic gain increased at a diminishing rate. The diminishing returns to selection show that the information content is not so large in the last phenotyped candidate compared to the first phenotyped candidate.

We tested for bias in predicted breeding values, as some concern has been raised that selective inclusion of animals in reference population may lead to biased

predictions (Boligon et al. (2012)). However, in no case was there detectable bias irrespective of criteria or proportion of phenotyped animals (results not shown). As expected, GBLUP model resulted to higher gains than BLUP (Figure 1). At 100% phenotyping the genetic gains from GBLUP were 15.5% higher than those from BLUP. The superiority of GBLUP could be attributed to higher accuracy (0.74) obtained relative to BLUP (0.60). This is consistent with previous studies which have demonstrated that, genomic selection is superior to traditional BLUP by 8-38% (Lillehammer et al. (2011)).

Conclusion

This study has shown that, it would be beneficial to rank selection candidates based on estimated breeding values and select the highest ranking individuals for phenotyping. Since there was no biasness detected, adoption of EBV-AP strategy will not only results to high genetic gains compared to random sampling, but will also require only phenotyping 80% of the available selection candidates to realize maximum achievable genetic gains in the breeding program. This could be important especially in terms of reducing production costs related to phenotyping in the breeding program. Our finding also suggests that, random sampling could be an alternative strategy but, more candidates will be required for phenotyping. Generally, irrespective of the strategy adopted there is diminishing return to selection as more candidates are phenotyped.

Literature Cited

- Boligon, A. A., Long, N., Albuquerque, L. G. et al. (2012). *J. Anim. Sci.*, 90:4716-4722.
- Gao, H., Christensen, O. F., Madsen, P. et al. (2012). *Genet. Sel. Evol.*, 44:8.
- Henryon, M., Berg, P., Ostensen, T. et al. (2012). *J. Animal Sci.*, 90:4681-4689.
- Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2011). *J. Anim. Sci.*, 89:3908-3916.
- Lillehammer, M., Meuwissen, T. H. E., and Sonesson, A. K. (2013). *Genet. Sel. Evol.*, 45:39.
- Pedersen, L. D., Sørensen, A. C., Henryon, M. et al. (2009). *Livest. Sci.*, 121:343-344.
- Wellmann, R., Preuß, S., Tholen, E. et al. (2013). *Genet. Sel. Evol.*, 45:28.
- VanRaden, P. M. (2008). *J. Dairy Sci.*, 91:4414-4423.