Accuracy of genomic estimated breeding values for crossbred body weight in broilers using a purebred or crossbred reference population.

P. Duenk[1], M.P.L. Calus[1], Y.C.J. Wientjes[1], V. Breen[2], J.M. Henshall[2], R. Hawken[2] & P. Bijma[1]

[1]Animal Breeding and Genomics, Wageningen University and Research, P.O. Box 338, 6700 AH Wageningen, the Netherlands
pascal.duenk@wur.nl (Corresponding Author)
[2] Cobb-vantress Inc., Siloam Springs, AR 72761-1030, USA

Summary

In breeding programs of pigs and poultry, selection is typically applied at the level of the purebred (PB) animals using PB data. This may result in a suboptimal response at the crossbred (CB) level when the genetic correlation between PB and CB performance () is lower than one. The objectives of this study, therefore, were to (1) estimate the , and (2) compare the accuracy of genomic estimated breeding values (GEBVs) for CB body weight in broilers using a PB reference population (RP) or a CB RP. Phenotype and genotype data were available for 5,274 PB and 10,395 CB offspring, and genotype data for their 163 PB sires. For comparison of cross-validation (CV) GEBV accuracies, the size of the CB RP was reduced to approximately match the size of the PB RP. Because the number of sires was limited, validation was also performed on GEBVs of CB offspring of the sires. The estimated was 0.94 (0.04). The results show that the accuracy of sire GEBVs was slightly higher with a CB RP (0.43) compared to a PB RP (0.40). Similarly, the accuracy of CB offspring GEBVs was higher with a CB RP (0.47) compared to a PB RP (0.25). These results are likely due to (1) the that is lower than one, and (2) differences in genomic relationships between reference and validation animals. In finalizing this work, we will develop an equation that uses the accuracy of CB GEBVs for CB performance to predict the accuracy of PB GEBVs for CB performance.

Keywords: genomic prediction, crossbreeding, crossbred performance, broilers

Introduction

In breeding programs of pigs and poultry, purebred (PB) animals coming from different lines are mated to produce crossbred (CB) production animals, because crossbreds tend to outperform their parents due to heterosis (Sellier 1976; Falconer and Mackay 1996). Although the aim of such breeding programs is to improve CB performance, selection is applied at the level of the purebreds, typically using information on PB performance. The response to selection at the CB level is therefore not maximized, because the genetic correlation between PB and CB performance () is generally lower than one (Wei and van der Werf 1995; Lukaszewicz et al. 2015; Wientjes and Calus 2017). Selection for CB performance requires methods to connect information from CB animals to PB selection candidates. To this end, genomic prediction methods may provide a solution.

In genomic prediction for CB performance, a CB reference population (RP) instead of a PB RP may be beneficial when (1) is low, (2) there are strong relationships between the CB

RP and selection candidates, and (3) the size of the CB RP is comparable to or greater than the size of the alternative PB RP (Dekkers 2007; Esfandyari et al. 2015; Van Grevenhof and Van Der Werf 2015; Hidalgo et al. 2016). In a study on pigs, Hidalgo et al. (2016) found that a CB RP did not increase the accuracy of genomic estimated breeding values (GEBVs) for CB performance compared to a PB RP. They concluded that this was due to low relationships between the CB animals and PB selection candidates, the high (~0.90) of the studied trait, and the much smaller size of the CB RPs (440 – 914) compared to the PB RPs (1292 – 1717) (Hidalgo et al. 2016). To our knowledge, there are no empirical studies that use a RP of CB broilers for genomic prediction of GEBVs for CB performance.

The objectives of this study were to (1) estimate the of body weight in broilers, and (2) compare the accuracy of genomic estimated breeding values (GEBVs) for CB body weight, using a CB RP or PB RP of equal size.

Materials and methods

Phenotypic data were available from 5,274 PB and 10,395 CB animals that were housed in the same environment. In total, these animals had 163 unique PB sires, of which 137 sires had both PB and CB offspring, 7 sires only had PB offspring, and 19 sires only had CB offspring. The trait analysed was body weight, measured between 33 to 36 days of age. Genotypes were available of all animals for 52,183 markers.

The was estimated with a bivariate model

, (1)

where is a vector of phenotypes, is a vector of fixed effects (breed, trial x pen, sex and age), is the design matrix of fixed effects, is a vector of genomic estimated breeding values (GEBVs) of all PB and CB offspring, is an incidence matrix linking breeding values to observations, and is a vector of random residuals. Subscripts denote if the terms relate to PB or CB data. The GEBVs were assumed to be , where () is the additive genetic variance in the purebreds (crossbreds), is the genetic covariance between PB and CB performance, and is the genomic relationship matrix of all PB and CB individuals. The -matrix was constructed with genotypes (0, 1, 2) and allele frequencies specific for the PB and CB animals, similar to the multi-population -matrix in Wientjes et al. (2017).

The second aim was to compare accuracies of sire GEBVs with a PB or CB RP of equal size. For each sire that had more than one CB offspring, half of its CB offspring were excluded. For sires that had only one CB offspring, no offspring were excluded. This procedure resulted in a CB RP of 5,156 animals. Subsequently, GEBVs were estimated with a univariate model

, (2)

where terms have the same meaning as in Equation 1, but with an overall mean instead of fixed effect for breed, because the phenotypes come from a single breed (PB or CB). Here, is a vector of GEBVs of the RP and validation animals, assumed to be , where is the additive genetic variance.

Accuracies were obtained using cross-validation (CV). To this end, sires were randomly assigned to one of five equally sized CV groups. For each CV group, GEBVs of

sires in that group were estimated using phenotypic data from offspring of sires in the other four CV groups. The random assignment to CV groups was repeated 10 times, such that for each scenario, 10 accuracies were obtained. Because the number of sires in each CV group was small (28 - 32), the standard error of the accuracy was expected to be relatively large (Daetwyler et al. 2013). Validation was therefore also performed at the level of the CB offspring of the sires, which allowed for larger CV groups, and thus smaller standard error of the accuracy.

The accuracy of sire GEBVs was calculated as the weighted correlation between GEBVs and the mean corrected phenotypes (MCP) of their CB offspring, divided by the square root of the weighted mean reliability of MCPs. Weights were derived using reliabilities of MCPs. The accuracy of CB offspring GEBVs was calculated as the correlation between GEBVs and their corrected phenotypes (CP), divided by the square root of the heritability in CB data.

All of the above analyses were carried out using the program MTG2 (Lee and van der Werf 2016).

Results and discussion

Genetic correlation

The estimated was 0.94 with a standard error of 0.04.

Accuracy of GEBVs

The accuracies of sire GEBVs and CB offspring GEBVs, using a PB or CB RP, are in Figure 1 (Appendix). The mean accuracy of sire GEBVs was a little higher with a CB RP (0.43) compared to a PB RP (0.40). The mean accuracy of CB offspring GEBVs was much higher with a CB RP (0.47) compared to a PB RP (0.25).

The estimated was lower than one, which resulted in higher accuracies with a CB RP than with a PB RP, for both sire GEBVs and CB offspring GEBVs. The difference in accuracy at the level of the CB offspring is, however, much larger than at the level of the sires. This is likely due to the fraction of genome shared between the reference and validation animals (i.e. genomic relationships). At the level of the CB offspring, the CB RP shares all haplotypes with the CB validation animals, which increases the CV accuracy. In contrast, the PB RP only shares half of its haplotypes with the CB validation animals, which decreases the CV accuracy. At the level of the sires, however, the CB RP shares only half of their haplotypes with the sire line, whereas the PB RP shares all haplotypes with the sire line. This favours the use of a PB RP at the sire level, but in this case it does not outweigh the benefit of a CB RP due to .

In a crossbreeding program, the selection candidates are PB animals. Hence, the accuracy of PB GEBVs for CB performance (PB-$GEBV_{CP}$) determines the response to selection and is the parameter of interest. Validation of PB-$GEBV_{CP}$ is, however, not always possible when for example phenotypes of CB relatives of selection candidates are not available. An alternative is to validate CB GEBVs for CB performance (CB-$GEBV_{CP}$) to estimate prediction accuracy (see for example Xiang et al. (2016)). Our results show, however, that validation of CB-$GEBV_{CP}$ overestimates the accuracy of PB-$GEBV_{CP}$, probably because the genomic relationships between the CB RP and CB validation animals are stronger than between the CB RP and PB validation animals. In the next step of this work, we will

investigate prediction of the accuracy of PB-GEBV$_{CP}$ using the accuracy of CB-GEBV$_{CP}$.

Acknowledgements

The authors thank the Netherlands Organisation of Scientific Research (NWO) and the Breed4Food consortium partners Cobb Europe, CRV, Hendrix Genetics, and Topigs Norsvin for their financial support, and Cobb Europe for providing the data.

List of References

Daetwyler, H.D., M.P. Calus, R. Pong-Wong, G. de Los Campos, and J.M. Hickey, 2013 Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. *Genetics* 193 (2):347-365.

Dekkers, J.C., 2007 Marker-assisted selection for commercial crossbred performance. *J Anim Sci* 85 (9):2104-2114.

Esfandyari, H., A.C. Sorensen, and P. Bijma, 2015 A crossbred reference population can improve the response to genomic selection for crossbred performance. *Genet Sel Evol* 47 (76).

Falconer, D., and T. Mackay, 1996 *Introduction to quantitative genetics*. Harlow: Pearson Education Limited.

Hidalgo, A.M., J.W. Bastiaansen, M.S. Lopes, M.P. Calus, and D.J. de Koning, 2016 Accuracy of genomic prediction of purebreds for cross bred performance in pigs. *J Anim Breed Genet* 133 (6):443-451.

Lee, S.H., and J.H.J. van der Werf, 2016 MTG2: an efficient algorithm for multivariate linear mixed model analysis based on genomic information. *Bioinformatics* 32 (9):1420-1422.

Lukaszewicz, M., R. Davis, J.K. Bertrand, I. Misztal, and S. Tsuruta, 2015 Correlations between purebred and crossbred body weight traits in Limousin and Limousin-Angus populations. *J Anim Sci* 93 (4):1490-1493.

Sellier, P., 1976 The basis of crossbreeding in pigs; A review. *Livest Prod Sci* 3 (3):203-226.

Van Grevenhof, I.E., and J.H. Van Der Werf, 2015 Design of reference populations for genomic selection in crossbreeding programs. *Genet Sel Evol* 47 (14).

Wei, M., and J.H. van der Werf, 1995 Genetic correlation and heritabilities for purebred and crossbred performance in poultry egg production traits. *J Anim Sci* 73 (8):2220-2226.

Wientjes, Y.C., P. Bijma, J. Vandenplas, and M.P.L. Calus, 2017 Properties of genomic relationships for estimating current genetic variances within and genetic correlations between populations. *bioRxiv*.

Wientjes, Y.C.J., and M.P.L. Calus, 2017 The purebred-crossbred correlation in pigs: a review of theory, estimates and implications. *J Anim Sci - in press*.

Xiang, T., B. Nielsen, G. Su, A. Legarra, and O.F. Christensen, 2016 Application of single-step genomic evaluation for crossbred performance in pig. *J Anim Sci* 94 (3):936-948.

Appendix

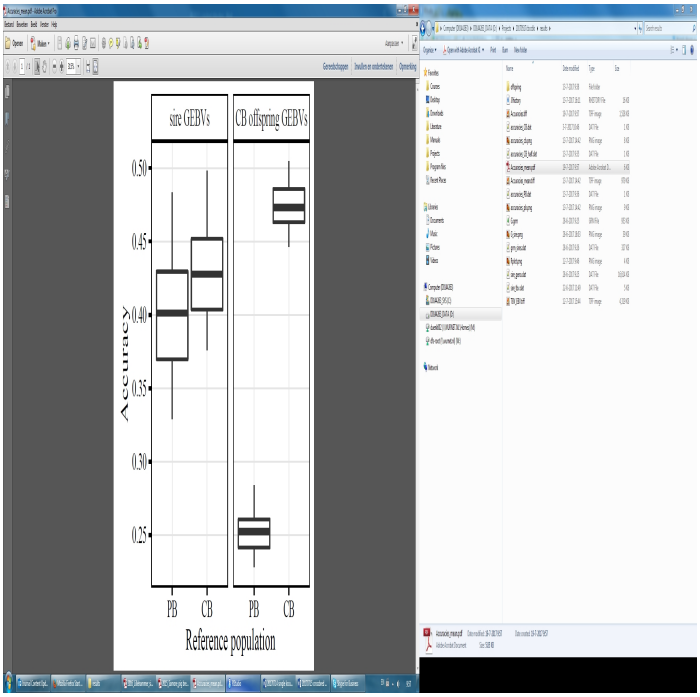

Figure 1. Accuracy of sire GEBVs (left) and CB offspring GEBVs (right) for CB body weight, using a purebred (PB) or crossbred (CB) reference population. Values are obtained from cross-validation. The boxes represent the mean (middle line), 25% percentile (bottom line) and 75% percentile (top line). Whiskers represent minimum and maximum values.