Breed-specific genome-wide association study for purebred and crossbred performance

C.A. Sevillano[1,2], S.E.F. Guimarães[3], F.F. Silva[3] & M.P.L. Calus[1]

[1] Wageningen University & Research Animal Breeding and Genomics, P.O. Box 338, 6700 AH Wageningen, The Netherlands
Claudia.sevillanodelaguila@wur.nl (Corresponding Author)
[2] Topigs Norsvin Research Center, P.O. Box 43, 6640 AA, Beuningen, The Netherlands
[3]Universidade Federal de Viçosa, Departmento de Zootecnia, Campus Universitario 36570.000, Viçosa, Minas Gerais, Brasil

Summary

The association of a given SNP with purebred (PB) performance might be different than its association with crossbred (CB) performance, because of interactions of the SNP marker with the environment or the genetic background. Therefore, the aim of the present study is to examine breed-specific associations of SNP to PB and CB performance. For this, we performed a genome-wide association study (GWAS) for back fat thickness (BF) with an approach that implements a genomic best linear unbiased prediction (GBLUP) model considering breed-of-origin of alleles. We observed some same regions for PB and CB performance, but the effects differed when observed in a PB or CB background. As expected, the breed with the lowest genetic correlation for BF between PB and CB, had fewest SNPs in common between PB and CB. Moreover, the effect of a given allele associated to BF in CB depended on the breed it was inherited from. These results suggest that SNP effects depend on the environment and on their genetic background, and are valuable to understand the low responses obtained when selecting PB animals for CB performance. The recognition of important regions associated to performance plus the differentiation of SNP effects according to their breed-of-origin, might inform future prediction models for CB performance.

Keywords: crossbred, back fat thickness, pigs, linkage disequilibrium, gwas

Introduction

Estimated SNP effects depend on the Linkage Disequilibrium (LD) between the SNP marker and the QTL, and on interactions of the SNP marker with the genetic background. When association analysis are performed on populations coming from breeding schemes with crosses, SNP by genetic background interactions may be more relevant, as the functional variation that underlies the QTL may not be segregating in one breed or epistatic interactions in one breed are different due to other genes that modify the effect of the QTL in that breed. With crossbreeding, SNP effects are observed in different environments and genetic backgrounds. Thus, the association of a given SNP to PB performance might differ from its association to CB performance, and an allele of that given SNP associated to CB performance could have a different effect depending on the breed it was inherited from, as it could be associated to a different underlying QTL for each breed.

Recently we have developed a procedure that enables breed-of-origin assignment of alleles in three-way CB animals (Vandenplas *et al.*, 2016, Sevillano *et al.*, 2016), which enables the use of a GBLUP model that accounts for breed-specific SNP effects (Sevillano *et*

al., 2017). With this tool, we can examine breed-specific associations of SNP to PB or CB performance. In the present study we aimed to performed a GWAS for BF, by estimating SNP effects from the solutions of a GBLUP considering breed-of-origin of alleles.

Material and methods

Data

The pig data consisted of three PB populations: 7547 Synthetic boar (S), 3288 Landrace (LR), and 12 794 Large White (LW) pigs. And 2816 three-way CB pigs (S (LR x LW) or S (LW x LR)). All pigs were genotyped using one of the three following SNP panels: Illumina PorcineSNP60.v2 BeadChip (60K.v2), Illumina PorcineSNP60 BeadChip (60K), or Illumina PorcineSNP10 BeadChip (10K). Pigs genotyped with the 60K or 10K chips were imputed to the 60K.v2 panel using FImpute Version 2.2 software (Sargolzaei *et al.*, 2014) following the same imputation strategy as Sevillano *et al.* (2016).

Phenotypes for BF (mm) were measured for PB using an ultrasound instrument at an average of age 173 days, while BF for CB were measured on the carcass using a probe, capteur gras maigre (CGM; Sydel, France) after slaughter when they achieved 120 kg (at an average age of 169 days).

Genome-wide association study

SNP allele substitution effects were estimated using BLUP similar to Wang *et al.* (2014), but instead of using a single-step BLUP, we used a GBLUP with breed-specific partial relationship matrices (Sevillano *et al.*, 2017). Briefly, SNP-allele effects were derived as follows:

1. Determine breed-of-origin of alleles to calculate breed-specific partial relationship matrices, $\mathbf{G}^{(S)}$, $\mathbf{G}^{(LR)}$, and $\mathbf{G}^{(LW)}$.
2. Calculate GEBVs for PB and CB performance using a GBLUP model with breed-specific partial relationship matrices.
3. Back-solve SNP-allele effects for PB and CB performance from GEBVs.
4. Calculate proportion of variance explained by non-overlapping blocks of SNPs.

Inference of the breed-of-origin of alleles.

To infer the breed-of-origin of the alleles in CB pigs we used the BOA approach developed by Vandenplas *et al.* (2016), with parameter settings recommended by Sevillano *et al.* (2016). The SNP set for this analysis consisted of SNPs that after imputation had a MAF >0.01 in the CB populations, and were segregating in all three PB populations (at least more than one observation of the minor allele). SNPs for which the paternal or maternal allele were assigned a breed-of-origin in less than 80% of the three-way CB pigs were removed. If an allele within a breed-of-origin was observed less than 6 times in the three-way CB population this SNP was also removed. The final SNP set for subsequent analyses consisted of 41,557 SNPs.

Model with three breed-specific partial relationship matrices.

To account for breed-specific effect of alleles, a 4-trait animal model with three breed-specific partial relationship matrices ($\mathbf{G}^{(S)}$, $\mathbf{G}^{(LR)}$, $\mathbf{G}^{(LW)}$) was fitted (BOA model, Sevillano *et al.*,

2017). Effects in the model included interaction between farm, breed and sex as fixed effects, and random common litter and additive genetic effects. Fixed effects also included body weight off-test for the PB, and trial and hot carcass weight for the CB.

The three breed-specific partial relationship matrices, $\mathbf{G^{(S)}}$, $\mathbf{G^{(LR)}}$, and $\mathbf{G^{(LW)}}$, were built using the breed-of-origin of phased alleles in CB pigs. The breed-specific partial relationship submatrices are defined as, e.g. for breed S origin,

$$\begin{aligned}\mathbf{G_{S,S}}&=(\mathbf{M^S}-2\mathbf{1p^{S\prime}})\mathbf{D^S}(\mathbf{M^S}-2\mathbf{1p^{S\prime}})'/N\\ \mathbf{G_{S,CB}}&=(\mathbf{M^S}-2\mathbf{1p^{S\prime}})\mathbf{D^S}(\mathbf{M^{CB}}-\mathbf{1p^{S\prime}})'/N\\ \mathbf{G_{CB,CB}}&=(\mathbf{M^{CB}}-\mathbf{1p^{S\prime}})\mathbf{D^S}(\mathbf{M^{CB}}-\mathbf{1p^{S\prime}})'/N\end{aligned} \quad (1)$$

where $\mathbf{M^S}$ is a matrix containing breed-specific allele content for S pigs (coded as 0, 1, or 2), $\mathbf{M^{CB}}$ is a matrix containing breed-specific allele content for CB pigs (coded as 0, or 1), vector $\mathbf{p^S}$ are breed S specific frequencies of the counted allele (p^S_j), calculated across S and CB pigs, $\mathbf{D^S}$ is diagonal with $D^S_{jj} = (2p^S_j(1- p^S_j))^{-1}$, and N is the number of SNPs. The breed-specific partial relationship submatrices $\mathbf{G^{(LR)}}$, and $\mathbf{G^{(LW)}}$ are defined similarly to $\mathbf{G^{(S)}}$. However the entries of $\mathbf{M^{CB}}$ matrix containing the breed-specific allele content for CB pigs are set to a missing value if the allele origin corresponds to the other maternal line and effectively did not contribute to the breed-specific partial relationship matrix. BLUP predictions of the BOA model were done using the MiXBLUP software (Ten Napel *et al.*, 2016). To estimate the variance components we used the same BOA model in the ASReml software (Gilmour *et al.*, 2009), after reducing each of the PB populations to 3500 animals most closely related to the CB population.

Backsolve SNP-allele effects from EBVs.

GEBVs of PB pigs for PB performance were converted to SNP-allele effects. For GEBVs of S pigs for PB performance ($\mathbf{a_S}$):

$$\begin{aligned}\mathbf{a_S} &= \mathbf{W^S b^S},\\ \mathbf{W^S} &= (\mathbf{M^S}-2\mathbf{1p^{S\prime}})\mathbf{U^S}\\ \mathbf{b^S} &= \mathbf{W^{S\prime}(W^S W^{S\prime})^{-1}a_S} = N^{-1}\mathbf{W^{S\prime}\,G^{(S)-1}a_S},\end{aligned} \quad (2)$$

where $\mathbf{W^S}$ contains centered genotypes, $\mathbf{b^S}$ are regression coefficients, $\mathbf{U^S}$ is an N x N diagonal matrix, with diagonal values of $(2p^S_j(1- p^S_j))^{-0.5}$. Note that $\mathbf{b^S}$ are not allele substitution effects, because $\mathbf{W^S}$ contain centered genotypes. The allele substitution effects, $\boldsymbol{\alpha_S}$, were obtained as $\mathbf{U^S b^S}$. The SNP-allele effects for CB performance, and for the other PB breeds were calculated similarly.

Variance proportion explained by SNP regions

To calculate the proportion of variance explained by a region, blocks of SNPs were built. Breakpoints between blocks of SNPs where defined when LD between two consecutive SNPs was not significant (*P*-value>0.05). LD and significance of LD was estimated by Fisher's exact test using the program Arlequin (Excoffier et al., 2005). Percentage of genetic variance explained by the *i*-th block of SNP was calculated as Wang *et al.* (2014), considering that a_i is the genetic value of the *i*-th block, σ^2_a is the total genetic variance, $\mathbf{z_j}$ is a vector of gene content of the *j*-th SNP for all PB individuals, and α_j is the effect of the *j*-th SNP within the *i*-

th block that contains n SNPs, and x_n is the mean length across blocks:

$$(\text{Var}(a_i) / \sigma^2_a) \times 100\% = (\text{Var}(\Sigma^n_{j=1} \mathbf{z}_j\alpha_j) / \sigma^2_a) \times (x_n/n) \times 100\% \quad (3)$$

Results and discussion

Variance components

The estimated heritability was larger for CB than for PB (0.41 for CB, vs. 0.31, 0.33, and 0.34 for S, LR, and LW, respectively). The genetic correlation between CB and PB performance was lowest in LR (0.71), followed by S (0.80), and LW (0.89).

Proportion of genetic variance explained by a region

The SNP blocks built based on alleles originated from the S paternal breed were longer than the blocks built based on alleles originated from the maternal breeds, as expected because the maternal haplotypes came from F1 pigs (*Table 1*). The proportion of genetic variance of BF explained by each SNP block in PB and CB of the different breeds are shown in *Figure 1.* The top 10 SNP blocks jointly explained 5.7%, 1.8 %, and 2.6% of the genetic variance for BF in S, LR, and LW, respectively. The top 10 SNP blocks jointly explained 4.9%, 2.4%, and 2.7% of the genetic variance for BF in CB when the alleles originate from S, LR, and LW, respectively. At the most, the top 10 SNP blocks covered 0.64% of the genome.

Table 1. Description of the blocks of SNPs built per breed.

		Length (# SNPs)		
Breed	*# of blocks*	*Mean*	*Min*	*Max*
S	5631	7.3	1	289
LR	7592	4.2	1	77
LW	6767	6.0	1	95

SNP effects associated to PB and CB performance

Among the top 10 SNP blocks for PB performance, there were 6, 4, and 5 in common with the top 10 blocks for CB performance, for breed S, LR, and LW, respectively. The LR breed had fewer blocks in common between PB and CB, as expected from the lower genetic correlation for BF between PB and CB for this breed. In general, we observed some same main SNPs across PB and CB performance, but we also observed that the effect of these SNP-alleles differed when observed in a PB or CB background. The differences in effects of SNP-allele can be due to genotyped by environment interactions (GxE) or differences in trait measurements between PB and CB (Wientjes & Calus, 2017; Godinho *et al.* 2017).

SNP effect associated to CB performance by breed-of-origin

As blocks across breed-of-origin were not the same, because of different patterns of LD, comparisons across breeds were done regarding the regions covered by the SNP blocks. Within the top 10, zero regions were in common across the 3 breeds-of origin, but 2 regions were in common across breeds-of origin S and LW. This indicates that for main regions, the

actual effect of a given SNP-allele associated to a performance in CB depended on the breed it was inherited from. Differences in effects of SNP-alleles across breed-of-origin can be due to different genetic background, i.e. epistasis and different allele frequencies. Main SNPs associated to BF are breed-of-origin specific, however, they explained a small part of the total variance, the large majority seem not to have a large effect on the trait and/or similar effect across breed-of-origin.

Conclusions

Some same main SNPs were observed across PB and CB performance but the effect of these SNP differed when observed in a PB or CB background. There were few regions in common across breeds, indicating that the actual effect of a given SNP-allele associated to a performance in CB depended on the breed it was inherited from. These results are valuable to understand the low responses obtained when selecting PB animals for CB performance. Moreover, the recognition of important regions associated to performance plus the differentiation of SNP-allele effects according to their breed-of-origin, might inform future prediction models for CB performance.

List of References

Excoffier, L., G. Laval & S. Schneider. Arlequin (version 3.0): an integrated software package for population genetics data analysis. Evolutionary bioinformatics online, 2005;1:47.

Gilmour A, B. Gogel, B. Cullis & R. Thompson, 2009. ASReml User Guide Release 3.0. VSN International Ltd, Hemel Hempstead, UK.

Godinho R. M., R. Bergsma, F. F. Silva, C. A. Sevillano, E. F. Knol, M. S. Lopes, P. S. Lopes, J. W. M. Bastiaansen & S. E. F. Guimarães, 2017. Genetic correlations between feed efficiency traits, and growth performance and carcass traits in purebred and crossbred pigs. J. Anim. Sc., Accepted.

Sargolzaei M., J.P.Chesnais & F.S. Schenkel, 2014. A new approach for efficient genotype imputation using information from relatives. BMC genomics. 15(1):1.

Sevillano C.A., Vandenplas J, Bastiaansen JWM, Calus MPL, 2016. Empirical determination of breed-of-origin of alleles in three-way crossbred pigs. Genet. Sel. Evol. 48(1): 55.

Sevillano C.A., Vandenplas J, Bastiaansen JWM, R. Bergsma & M.P.L. Calus, 2017. Genomic evaluation for a three-way crossbreeding system considering breed-of-origin of alleles. Genet. Sel. Evol. 49(1):75.

Ten Napel J, M.P.L. Calus, M. Lidauer, I. Stranden, E. Mäntysaari, H. Mulder & R. Veerkamp, 2016. MiXBLUP, the Mixed-model Best Linear Unbiased Prediction software for PCs for large genetic evaluation systems. Version 2.0, Wageningen, the Netherlands.

Vandenplas J., M.P.L. Calus, C.A. Sevillano, J.J. Windig & J.W.M. Bastiaansen, 2016. Assigning breed origin to alleles in crossbred animals. Genet. Sel. Evol. 48(1): 61.

Wang, H., I. Misztal, , I. Aguilar, A. Legarra, R.L. Fernando, Z. Vitezica, R. Okimoto, T. Wing, R. Hawken & W. M. Muir, 2014. Genome-wide association mapping including phenotypes from relatives without genotypes in a single-step (ssGWAS) for 6-week body weight in broiler chickens. Front. Genet. 5: 134.

Wientjes Y & Calus M, 2017. BOARD INVITED REVIEW: The purebred-crossbred correlation in pigs: A review of theory, estimates, and implications. J. Anim. Sc., 95(8):3467-3478.

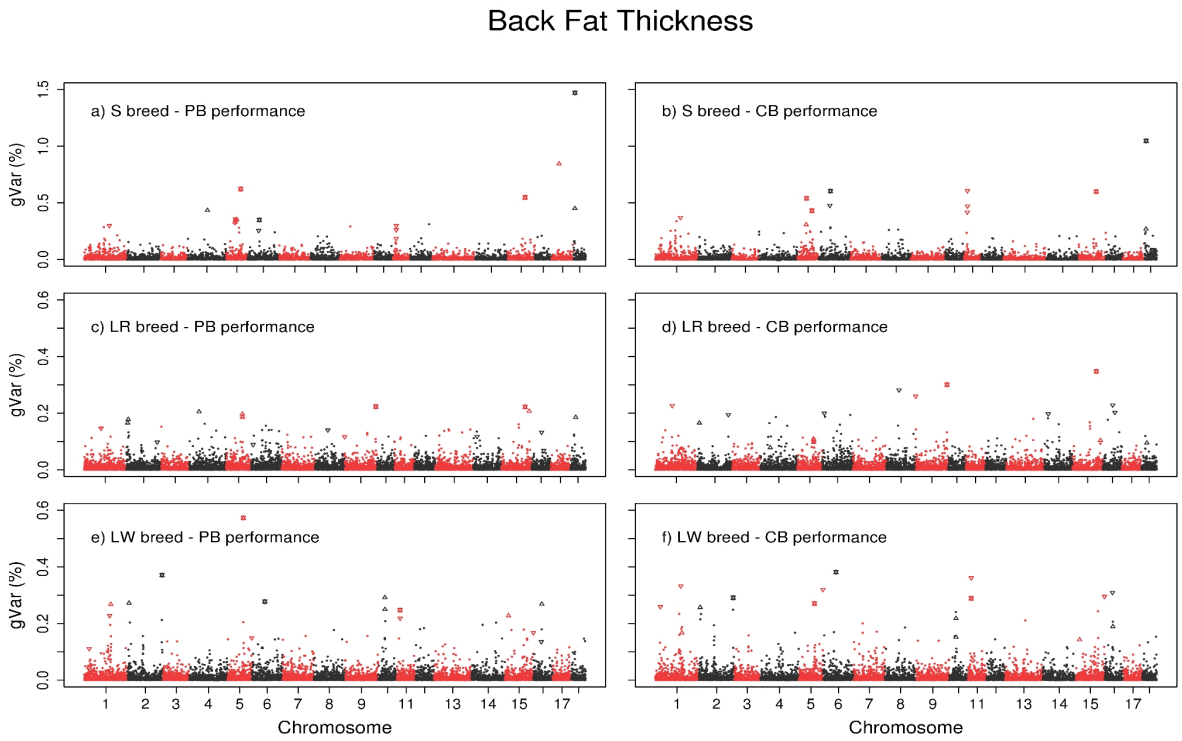

Figure 1. Proportion of genetic variance for back fat thickness explained by each SNP block of SNPs in Synthetic boar (S breed) pigs (a), in Landrace (LR breed) pigs (c), in Large White (LW breed) pigs (e), and in crossbred pigs (CB) when alleles originate from S breed (b), LR breed (d), or LW breed (f). Top 10 SNP blocks explaining the variance for PB (△), and the top 10 SNP blocks explaining the variance for CB performance (▽). SNP blocks belonging to the top 10 in both, PB and CB performance (✡)