Using SNP Markers to Estimate Additive, Dominance and Imprinting Genetic Variance

M.S. *Lopes*[*,§], J.W.M. Bastiaansen[§], L. Janss[‡], H. Bovenhuis[§] and E.F. Knol[*].
[*]TOPIGS Research Center IPG, Beuningen, the Netherlands, [§]Animal Breeding and Genomics Centre, Wageningen University, Wageningen, the Netherlands, [‡] Centre for Quantitative Genetics and Genomics, Aarhus University, Tjele, Denmark

ABSTRACT: The contributions of additive, dominance and imprinting effects to the variance of number of teats (NT) were evaluated in two purebred pig populations using SNP markers. Three different random regression models were evaluated, accounting for the mean and: 1) additive effects (MA), 2) additive and dominance effects (MAD) and 3) additive, dominance and imprinting effects (MADI). Additive heritability estimates were 0.30, 0.28 and 0.27-0.28 in both lines using MA, MAD and MADI, respectively. Dominance heritability ranged from 0.06 to 0.08 using MAD and MADI. Imprinting heritability ranged from 0.01 to 0.02. Dominance effects make an important contribution to the genetic variation of NT in the two lines evaluated. Imprinting effects appeared less important for NT than additive and dominance effects. The SNP random regression model presented and evaluated in this study is a feasible approach to estimate additive, dominance and imprinting variance.
Keywords: pig; teat number, dominance; imprinting

Introduction

In pigs and other livestock species, selection has been carried out in purebred populations although the final goal is to improve crossbred performance. Crossbreeding takes advantage of breed complementarity and non-additive effects. Thus, to produce a better crossbred animal, it is necessary to select purebred animals in order to produce favorable allelic combinations in the crossbred generation (Zeng et al. (2013)). Including dominance and imprinting effects in genetic evaluations of crossbreeding schemes may, therefore, lead to a better crossbred product.

Traditional and genomic selection has mainly focused on additive genetic effects. Previous studies have used pedigree information to estimate dominance (de Boer et al. (1993); Misztal (1997)) and imprinting effects (de Vries et al. (1994); Tier and Meyer (2012)), but using pedigree based methods it is difficult to accurately estimate dominance effects and to disentangle imprinting, maternal and permanent environmental effects (Tier and Meyer (2012)). Although the availability of genomic information has increased the possibilities to quantify dominance and imprinting effects, only recently attempts have been made to quantify and to exploit the proportion of genetic variance due to dominance effects (Toro and Varona (2010); Su et al. (2012); Zeng et al. (2013); Da et al. (2014)). The contribution of imprinting effects to the total genetic variance, however, has not been investigated using genomic data.

In this study, the trait 'number of teats' (NT) was studied because QTLs with dominance and imprinting effects have previously been reported for this trait (Hirooka et al. (2001); Ding et al. (2009)). In addition, NT can be measured in males and females, which increases the number of animals with proper phasing and with phenotypic records. The objective of this study was to estimate the contribution of additive, dominance and imprinting effects to the variance of NT in two different purebred pig populations.

Materials and Methods

Data. A total of 2,013 Landrace-based (LN) and 2,402 Large White-based (LW) animals, genotyped using the Illumina Porcine SNP60 Beadchip (Ramos et al. (2009)) were available for this study. All animals had a frequency of missing genotypes <0.05. SNPs with call rate <0.95, minor allele frequency <0.01, strong deviation from Hardy-Weinberg equilibrium (χ^2>600), GenCall<0.15, unmapped SNPs and SNPs located on either the X or Y chromosome, according to the Sscrofa10.2 assembly of the reference genome (Groenen et al. (2012)), were excluded from the data set. After quality control, 39,089 SNPs for LN and 37,796 SNPs for LW out of the initial 64,232 SNPs were kept for phasing procedures.

The phasing and imputation of missing genotypes were performed within line using AlphaImpute (Hickey et. al (2011)), which combines genomic and pedigree information to determine the parental origin of alleles. The pedigree depth used in this analysis was up to 5 generations (between genotyped animals). To ensure accurate phasing, only animals which had both parents or at least one parent and one sib genotyped were used. Due to these restrictions, 1,538 LN and 2,112 LW animals were used for the estimation of variance components. All these animals had their individual NT counted at birth. The average NT in the LN population was 15.62±1.04, ranging from 12 to 20 teats. In the LW population the average NT was 15.37±0.97, ranging from 14 to 20 teats.

Statistical analyses. Parameters were estimated using a random regression model implemented in the program BayZ (http://www.bayz.biz/). The analyses were performed within line using three different models:

$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Xb} + \mathbf{Aa} + \mathbf{e} \quad \text{(MA)}$$
$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Xb} + \mathbf{Aa} + \mathbf{Dd} + \mathbf{e} \quad \text{(MAD)}$$
$$\mathbf{y} = \mathbf{1}\mu + \mathbf{Xb} + \mathbf{Aa} + \mathbf{Dd} + \mathbf{Ii} + \mathbf{e} \quad \text{(MADI)}$$

where $\mathbf{y}$ is a vector of phenotypic observations; μ is the mean of the populations and $\mathbf{1}$ a vector of ones; $\mathbf{X}$ is the design matrix for the fixed effects (sex and herd-year-season of birth); $\mathbf{b}$ is a unknown vector of fixed effects; $\mathbf{A}$, $\mathbf{D}$ and $\mathbf{I}$ are design matrices with regressors for additive, dominance and imprinting effects, respectively; $\mathbf{a}$, $\mathbf{d}$ and $\mathbf{i}$ are unknown vectors of additive, dominance and imprinting effects, respectively; and $\mathbf{e}$ is a vector of residuals. The

entries of the design matrices **A**, **D** and **I** are regressors calculated from the observed phased probabilities of the marker genotypes. For each SNP of each animal, AlphaImpute (Hickey et. al (2011)) generates two probabilities: P_1 being the probability that a specific allele was received from its father, say an allele G of a GC SNP, and P_2 the probability that a G allele was received from its mother. Considering a heterozygous animal (GC) where the G allele was inherited, with certainty, from its father (and therefore a C allele from its mother), the probabilities would be P_1=1 and P_2=0. In order to obtain the regressors that constitute the entries of **A** (regA), **D** (regD) and **I** (regI), the following transformation of these probabilities was applied:

$$\text{regA} = P_1 + P_2 \qquad \text{regD} = |P_1 - P_2| \qquad \text{regI} = P_1 - P_2$$

thus, the genotypes (GG, GC, CG, GG) were recoded as (0, 1, 1, 2) for **A**, (0, 1, 1, 0) for **D** and (0, -1, 1, 0) for **I**. In cases where the phasing could not be established with certainty, the regressors were based on the probabilities.

Uniform distributions were assigned to the variances. Assumed distributions of SNP effects were: $\mathbf{a} \sim N(0,\sigma_a^2)$, $\mathbf{d} \sim N(0,\sigma_d^2)$, $\mathbf{i} \sim N(0,\sigma_i^2)$ and $\mathbf{e} \sim N(0,\sigma_e^2)$, with σ_a^2, σ_d^2, σ_i^2 and σ_e^2 being the additive, dominance, imprinting and residual variance, respectively. The total variance in the model can be expressed as $(\sigma^2_{\mathbf{Aa}} + \sigma_e^2)$ for MA, $(\sigma^2_{\mathbf{Aa}} + \sigma^2_{\mathbf{Dd}} + \sigma_e^2)$ for MAD and $(\sigma^2_{\mathbf{Aa}} + \sigma^2_{\mathbf{Dd}} + \sigma^2_{\mathbf{Ii}} + \sigma_e^2)$ for MADI, being $\sigma^2_{\mathbf{Aa}}$=**AA'**σ_a^2, $\sigma^2_{\mathbf{Dd}}$=**DD'**σ_d^2 and $\sigma^2_{\mathbf{Ii}}$=**II'**σ_i^2. Alternatively, the total genomic variance explained by additive ($\sigma^2_{\mathbf{Aa}}$), dominance ($\sigma^2_{\mathbf{Dd}}$) and imprinting ($\sigma^2_{\mathbf{Ii}}$) effects can be obtained by evaluating *var*(**Aa**), *var*(**Dd**) and *var*(**Ii**), respectively, per MCMC cycle. This allowed to obtain posterior means and posterior standard deviation (SD) for all parameters.

Each model was run as a single chain with a length of 1,000,000 which was sampled each 100 iterations. The first 15,000 iterations were regarded as burn-in period. Therefore, results from 9,850 samples were used to estimate the posterior mean of the parameters. As part of the diagnostics analyses of the MCMC chain, the R package CODA (Plummer et al. (2006)) was used to estimate the standard error (SE) of the chain, which was defined as $(\sigma_p^2 / \sqrt{ESS})$, where σ_p^2 was the variance of the p^{th} parameter and *ESS* was the effective sample size after correction for autocorrelation across MCMC samples (Plummer et al. (2006)). Following the Monte Carlo theory, the asymptotic 95% confidence interval (CI= $\sigma_p^2 \pm 1.96$*SE) was estimated.

Results and Discussion

The proportions of phenotypic variance explained by all genetic effects evaluated in this study were very similar between the two populations (Table 1). All parameters were statistically different from zero (based on the asymptotic 95% CI).

Additive. Using the MA model, the estimated heritability for NT was 0.30 (Table 1) in both populations, which was slightly lower than the pedigree-based estimates (0.31±0.07 for LN and 0.32±0.05 for LW) using the same data. In both lines, the additive heritability estimate was 0.28 using the MAD model. Using the MADI model, the additive heritability was slightly different between the two lines (0.27 for LN and 0.28 for LW). Using the MADI model compared to the MA model, the total heritability of NT (sum of the proportion of phenotypic variance explained by all genetic effects included in the model) increased by 23%, from 0.298 to 0.368, in the LN and by 21%, from 0.303 to 0.368, in the LW population. A reduction of additive genetic variance and an increase (27%) of the total heritability was also observed by Su et al. (2012) when non-additive genetic effects (dominance and epistasis) were included in the model to evaluate daily gain in pigs. This reduction in the additive genetic variance was expected, since dominance effects also contribute to the additive genetic variance in the MA model (Falconer and Mackay (1996)). Therefore, applying a model that accounts for additive and dominance effects separately may result in dominance effects contributing less to the additive genetic variance.

Table 1. Estimated variance components and heritabilities. Total additive, dominance, imprinting and residual variances (σ^2_{Aa}, σ^2_{Dd}, σ^2_{Ii} and σ_e^2, respectively) and additive, dominance and imprinting heritabilities (h_a^2, h_d^2 and h_i^2, respectively) of number of teats in two purebred pig population using three different models.

	Landrace			Large-white		
	PM[†]	**SD**[¥]	**SE**[&]	**PM**	**SD**	**SE**
MA[*]						
σ_e^2	0.730	0.047	0.001	0.623	0.029	0.001
σ^2_{Aa}	0.310	0.073	0.002	0.272	0.032	0.001
h_a^2	0.298	0.055	0.002	0.303	0.031	0.001
MAD[**]						
σ_e^2	0.692	0.055	0.003	0.582	0.036	0.002
σ^2_{Aa}	0.294	0.072	0.002	0.252	0.035	0.001
σ^2_{Dd}	0.061	0.050	0.007	0.064	0.035	0.004
h_a^2	0.279	0.056	0.002	0.280	0.035	0.001
h_d^2	0.058	0.048	0.007	0.071	0.039	0.004
MADI[***]						
σ_e^2	0.662	0.071	0.005	0.568	0.036	0.002
σ^2_{Aa}	0.289	0.096	0.006	0.249	0.037	0.001
σ^2_{Dd}	0.083	0.054	0.006	0.072	0.036	0.004
σ^2_{Ii}	0.021	0.039	0.003	0.011	0.009	0.001
h_a^2	0.271	0.058	0.003	0.276	0.037	0.002
h_d^2	0.078	0.045	0.005	0.080	0.040	0.004
h_i^2	0.019	0.022	0.003	0.012	0.010	0.001

[†]PM: posterior mean;
[¥]SD: standard deviation of the posterior mean;
[&]SE: standard error of the MCMC;
[*]MA: model that includes only additive genetic effects;
[**]MAD: model that includes additive and dominance genetic effects;
[***]MADI: model that includes additive, dominance and imprinting genetic effects.

Dominance. The proportion of phenotypic variance explained by dominance effects ranged between 6 and 8% using MAD and MADI models for both lines. The estimated dominance variance in proportion to the additive variance was 21% for LN and 24% for LW using the MAD model. These results indicate that dominance effects make an important contribution to the genetic variation of NT. For daily gain in pigs, Su et al. (2012) estimated that dominance effects accounted for 5.6% of the total phenotypic variance (26% in proportion to additive effects). Based on pedigree estimates, Culbertson et al. (1998) reported that the ratio of dominance to additive variances for different traits in pigs ranged from 11 to 78%. Interestingly, there was an increase in dominance variance when imprinting effects (MADI) where included in the model (from 0.061 to 0.083 for LN and from 0.064 to 0.072 for LW), although the SD of the posterior mean remained approximately the same for both lines (Table 1).

In recent studies, dominance effects have been evaluated in genomic prediction scenarios. Including dominance effects in genomic evaluations has been reported to increase the accuracy and decreased the bias of estimated breeding values (Toro and Varona (2010); Su et al. (2012)). In addition, using dominance in genomic evaluations is expected to result in greater cumulative response to selection of purebred animals for crossbred performance than additive models, especially in the presence of overdominance and when retraining is not performed at each generation (Zeng et al. (2013)). Even when purely additive effects were evaluated, the inclusion of dominance in the genomic evaluations did not result in a decrease in accuracy of prediction (Toro and Varona (2010); Su et al. (2012); Zeng et al. (2013)).

Imprinting. Imprinting effects accounted for a small proportion of the phenotypic variance (1-2%) in the two lines (Table 1). The posterior mean and SD for imprinting effects were 0.021±0.039 for LN and 0.011±0.009 for LW. Although the SD were high, the asymptotic 95% CI ranged from 0.015 to 0.027 for LN and from 0.009 to 0.013 for LW, showing that the estimates of imprinting variance was statistically different from zero. The presence of imprinting variance is consistent with two imprinted QTL reported by Hirooka et al. (2001) on chromosome 2 and 12. However these two QTL alone explained 1.3 and 2.2% of the phenotypic variance. These larger imprinting variances may in part be explained by the design, an experimental F2 population, analysed in the QTL study.

Due to the low proportion of phenotypic variance explained by imprinting variance, imprinting effects may not be highly relevant in selection for NT. However, this study shows that is feasible to estimate imprinting variance with the presented method, which can be used to evaluate the relevance of imprinting for other traits. Combining imprinting and dominance effects with mate allocation techniques, under a genomic selection scenario, opens new perspectives for the optimization of breeding programs aiming for an improved performance of crossbred animals.

Conclusion

Dominance effects make an important contribution to the genetic variation of NT in the two lines evaluated. Although imprinting effects may not be highly relevant for NT, it can be concluded that the method used in this study allows estimating additive, dominance and imprinting variance. These results open new perspectives for the inclusion of dominance and imprinting effects in genetic evaluations, especially regarding mate allocation techniques for the optimization of crossbreeding programs. The predictability of an individual's total genetic merit using additive, dominance and imprinting simultaneously needs to be further evaluated.

Literature Cited

Culbertson, M. S., Mabry, J. W., Misztal, I. et al. (1998). *J. Anim. Sci.* 76: 448-451.

Da, Y., Wang, C., Wang, S. et al. (2014). *PLoS One.* 9: e87666.

De Boer, I. J. M., and Hoeschele, I. (1993). *Theor. Appl. Genet.* 86: 245-258.

De Vries, A. G., Kerr, R., Tier, B. et al. (1994). *Theor. Appl. Genet.* 88:1037-1042.

Ding, N., Guo, Y., Knorr, C. et al. (2009). *BMC Genet.*, 10:6.

Falconer D. S. and Mackay T. F. C. (1996). Harlow: Longmans Green.

Groenen M. A. M., Archibald A. L., Uenishi H. et al. (2012). *Nature.* 491:393-398.

Hickey, J. M., Kinghorn, B. P., Tier, B. et al. (2011). *Genet. Sel. Evol.* 43:12.

Hirooka, H., De Koning, D. J., Harlizius, B. et al. (2001). *J. Anim. Sci.* 79:2320-2326.

Misztal, I. (1997). J. Dairy Sci. 80: 965-974.

Plummer, M., Best, N., Cowles K. et al. (2006). *R News*. 6:7-11.

Ramos A. M., Crooijmans R. P. M. A., Affara N. A. et al. (2009). *PLoS One*. 4:1-13.

Su, G., Christensen, O. F., Ostersen, T. et al. (2012). *PLoS One*. 7:e45293.

Tier, B., and Meyer, K. (2012). *J. Anim. Breed. Genet.* 5: 359-368.

Toro, M. A., and Varona, L. (2010). *Genet. Sel. Evol.* 42: 33.

Zeng, J., Toosi, A., Fernando, R. L. et al. (2013). *Genet. Sel. Evol.* 45:11.