



**HAL**  
open science

## Dominance and epistatic genetic variances for litter size in pigs using genomic models

Zulma G. Vitezica, Antonio Reverter, William Herring, Andres Legarra

► **To cite this version:**

Zulma G. Vitezica, Antonio Reverter, William Herring, Andres Legarra. Dominance and epistatic genetic variances for litter size in pigs using genomic models. *Genetics Selection Evolution*, 2018, 50 (1), pp.71. 10.1186/s12711-018-0437-3 . hal-02405297

**HAL Id: hal-02405297**

**<https://hal.science/hal-02405297>**

Submitted on 11 Dec 2019

**HAL** is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution 4.0 International License

RESEARCH ARTICLE

Open Access



# Dominance and epistatic genetic variances for litter size in pigs using genomic models

Zulma G. Vitezica<sup>1\*</sup> , Antonio Reverter<sup>2</sup>, William Herring<sup>3</sup> and Andres Legarra<sup>4</sup>

## Abstract

**Background:** Epistatic genomic relationship matrices for interactions of any-order can be constructed using the Hadamard products of orthogonal additive and dominance genomic relationship matrices and standardization based on the trace of the resulting matrices. Variance components for litter size in pigs were estimated by Bayesian methods for five nested models with additive, dominance, and pairwise epistatic effects in a pig dataset, and including genomic inbreeding as a covariate.

**Results:** Estimates of additive and non-additive (dominance and epistatic) variance components were obtained for litter size. The variance component estimates were empirically orthogonal, i.e. they did not change when fitting increasingly complex models. Most of the genetic variance was captured by non-epistatic effects, as expected. In the full model, estimates of dominance and total epistatic variances (additive-by-additive plus additive-by-dominance plus dominance-by-dominance), expressed as a proportion of the total phenotypic variance, were equal to 0.02 and 0.04, respectively. The estimate of broad-sense heritability for litter size (0.15) was almost twice that of the narrow-sense heritability (0.09). Ignoring inbreeding depression yielded upward biased estimates of dominance variance, while estimates of epistatic variances were only slightly affected.

**Conclusions:** Epistatic variance components can be easily computed using genomic relationship matrices. Correct orthogonal definition of the relationship matrices resulted in orthogonal partition of genetic variance into additive, dominance, and epistatic components, but obtaining accurate variance component estimates remains an issue. Genomic models that include non-additive effects must also consider inbreeding depression in order to avoid upward bias of estimates of dominance variance. Inclusion of epistasis did not improve the accuracy of prediction of breeding values.

## Background

Genomics provides tools to understand the effects of genes and their interactions and new approaches for genetic improvement [1]. In quantitative genetics, partitioning genetic variance for a trait into statistical components due to additivity, dominance, and epistasis is useful for prediction and selection, even if it does not reflect the biological (or functional) effect of the underlying genes [2]. Additive, dominance and epistatic genetic variances

of quantitative traits are required to estimate breeding values and for making optimal selection decisions.

Within-breed non-additive effects, and in particular epistasis, are often ignored in genetic improvement programs. However, the total genetic value of an animal is a function of both additive and non-additive effects and, taken together, these effects could result in better predictors of future phenotypes [3] and inform mate allocation. Indeed, assortative mating can improve the performance of livestock when dominance [4] and/or epistasis are/is present [5]. Evidence of non-additive variance in commercially important traits (e.g. for body depth in fish, [6]) has opened opportunities for specialized breeding schemes. To take the effects of dominance and/or

\*Correspondence: zulma.vitezica@ensat.fr

<sup>1</sup> INP ENSAT, UMR 1388 GenPhySE, 31326 Castanet-Tolosan, France  
Full list of author information is available at the end of the article



epistasis into account, it is necessary to refine the estimation of non-additive variance components from phenotypic data.

The additive genetic effect is the part of an individual's total genetic effect that is transmissible across generations from parents to offspring. In contrast, non-additive genetic effects (dominance and epistasis) can be seen as the residual genetic effect after fitting additive substitution effects. Although most of the genetic variance is additive [1, 7] and captures most of the functional epistatic action of genes, the epistatic variance should not be neglected. Knowing its magnitude in real data and exploring the predictive ability of a model that accounts for epistatic effects are of great relevance. It is also of interest to know how much gene-by-gene ( $G \times G$ ) interaction exists (e.g. to determine the effect of the same allele in different breeds) or to account for the contribution of epistasis to the creation of "new" additive variance over time [8].

Several authors have proposed the inclusion of dominance and epistatic effects in genetic evaluation using high-density marker genotypes (genomic evaluation) [5, 9–12]. Most epistatic models consider only additive-by-additive epistatic interactions (e.g. [9, 12]), although dominance-by-dominance and dominance-by-additive interactions may play a major role in heterosis and in inbreeding and outbreeding depression (see [13] p. 223). Recently, Vitezica et al. [14] proposed a flexible and general approach to construct "genomic" relationship matrices for populations that are or are not in Hardy–Weinberg equilibrium (HWE), e.g. F1 crosses. They proved that epistatic genomic relationship matrices for two or higher order interactions can be constructed using Hadamard products of additive and dominance genomic orthogonal relationships, regardless of the existence of HWE. They also pointed out that, nevertheless, standardization of genomic relationship matrices based on the trace of the relationship matrices is needed. However to date, models that use these relationship matrices have not been applied to real data.

Xiang et al. [15] proved analytically that, in the presence of directional dominance, inclusion of genomic inbreeding as a covariate is necessary to obtain correct estimates of dominance variance. Genomic inbreeding of an individual was shown to be correctly defined as the proportion of genotyped single nucleotide polymorphisms (SNPs) at which the individual is homozygous [16]. This was confirmed in real data by Xiang et al. [15] and Aliloo et al. [4]. The effect of fitting or not fitting genomic inbreeding on estimates of epistatic variance components is unknown.

In current pig production systems, the number of pigs weaned is a key factor to increase productivity, and litter

size (e.g., total number of piglets born per litter) is one of the most important traits under selection, in maternal line breeding programs [17]. Although litter size in pigs has a low narrow-sense heritability [17–19], non-additive variance could still be abundant and should be ascertained. This is particularly important because the estimation of non-additive effects (dominance and epistasis) for litter size is of increasing interest since it can be used to define mate allocation strategies between selection candidates for developing new crossbreeding or even purebred breeding schemes.

The objectives of this work were to estimate additive and non-additive (dominance and epistasis) variance components for litter size for a real pig population, by considering or not inbreeding depression and to determine whether the accuracy of prediction of breeding values and total genetic values increases with the inclusion of dominance and epistatic effects.

## Methods

### Phenotypic and genomic data

Data for this study were provided by Genus plc (Hendersonville, TN, USA). Animal Care and Use Committee approval was not necessary for this study because the data were obtained from an existing database. Data on litter size (total number of piglets born per litter) were from a pig pure line. The average litter size was equal to  $12.7 \pm 3.1$  and 13,369 records were available for 3619 sows. Genotypes for all sows were generated using the Illumina PorcineSNP60 BeadChip (Illumina, San Diego, CA). After quality control using default parameters by preGSf90 [20] (HWE, minor allele frequency, SNP call rate and animal call rate), 38,779 autosomal SNPs remained and were used to build genomic relationship matrices.

### Genomic evaluation models

Phenotypes were analyzed using a genomic best linear unbiased prediction (GBLUP) (mixed) model. Parity number and the combined effect farm-year-month of farrowing were included as fixed effects. The model also included a random permanent environmental effect for each sow. The linear model including additive, dominant and interaction terms can be written as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{f}b + \mathbf{Z}\mathbf{g}_A + \mathbf{Z}\mathbf{g}_D + \mathbf{Z} \sum_{i=A,D} \sum_{\substack{j=A,D \\ i \geq j}} \mathbf{g}_{ij} + \mathbf{Z}\mathbf{p}\mathbf{e} + \mathbf{e},$$

where  $\mathbf{y}$  is the vector of phenotypic records,  $\boldsymbol{\beta}$  is the fixed effect vector,  $\mathbf{f}b$  models the inbreeding depression, where  $\mathbf{f}$  is the vector of genomic inbreeding coefficients based on the proportion of homozygous SNPs and  $b$  is the inbreeding depression parameter per unit of inbreeding,  $\mathbf{g}_A$  is a vector of breeding values of the sows,  $\mathbf{g}_D$  is a

vector of dominance deviations,  $\mathbf{g}_{ij}$  is a vector of epistatic genetic values,  $\mathbf{pe}$  is the permanent environmental effect vector ( $Var(\mathbf{pe}) = \mathbf{I}\sigma_{pe}^2$ ),  $\mathbf{e}$  is a residual vector ( $Var(\mathbf{e}) = \mathbf{I}\sigma_e^2$ ), and  $\mathbf{X}$  and  $\mathbf{Z}$  are design matrices relating records to fixed effects and genetic and permanent environmental effects, respectively. The epistatic genetic effects can be partitioned into additive-by-additive ( $\mathbf{g}_{AA}$ ), additive-by-dominance ( $\mathbf{g}_{AD}$ ), and dominance-by-dominance ( $\mathbf{g}_{DD}$ ) effects. Higher order epistasis interactions were not considered as they were deemed to be negligible [21] or too difficult to estimate. Covariance matrices of genetic effects were as follows:

$$\begin{aligned} Var(\mathbf{g}_A) &= \mathbf{G}_A\sigma_A^2, & Var(\mathbf{g}_D) &= \mathbf{G}_D\sigma_D^2, \\ Var(\mathbf{g}_{AA}) &= \mathbf{G}_{AA}\sigma_{AA}^2, & Var(\mathbf{g}_{AD}) &= \mathbf{G}_{AD}\sigma_{AD}^2, \\ Var(\mathbf{g}_{DD}) &= \mathbf{G}_{DD}\sigma_{DD}^2 \end{aligned}$$

where the covariance matrices  $\mathbf{G}_A$  and  $\mathbf{G}_D$  for additive and dominance effects, which are involved in the construction of covariance matrices  $\mathbf{G}_{AA}$ ,  $\mathbf{G}_{AD}$  and  $\mathbf{G}_{DD}$  for epistatic effects, were constructed assuming HWE. The additive relationship matrix was calculated as in VanRaden [22]:  $\mathbf{G}_A = \frac{\mathbf{MM}'}{2\sum_i p_i q_i}$ , where matrix  $\mathbf{M}$  has elements that are equal to  $(2 - 2p_i)$ ,  $(1 - 2p_i)$ ,  $-2p_i$  for genotypes  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$ , respectively, where  $p_i$  is the frequency of allele  $A_1$  at SNP  $i$ . The dominance relationship matrix was computed as  $\mathbf{G}_D = \frac{\mathbf{WW}'}{4\sum_i p_i^2 q_i^2}$  where  $\mathbf{W}$  has elements equal to  $-2q_i^2$ ,  $2p_i q_i$ ,  $-2p_i^2$  for genotypes  $A_1A_1$ ,  $A_1A_2$  and  $A_2A_2$ , respectively [10]. Covariance matrices for epistatic effects were computed using the Hadamard products and traces as  $\mathbf{G}_{AA} = \frac{\mathbf{G}_A \odot \mathbf{G}_A}{tr(\mathbf{G}_A \odot \mathbf{G}_A)/n}$ ,  $\mathbf{G}_{AD} = \frac{\mathbf{G}_A \odot \mathbf{G}_D}{tr(\mathbf{G}_A \odot \mathbf{G}_D)/n}$ , and  $\mathbf{G}_{DD} = \frac{\mathbf{G}_D \odot \mathbf{G}_D}{tr(\mathbf{G}_D \odot \mathbf{G}_D)/n}$  [14].

In these analyses, we implicitly assume that allele substitution effects of quantitative trait loci (QTL) are distributed as  $\alpha \sim N(0, \mathbf{I}\sigma_\alpha^2)$ , dominance effects as  $\mathbf{d} \sim N(0, \mathbf{I}\sigma_d^2)$ , additive-by-additive epistatic effects as  $(\alpha\alpha) \sim N(0, \mathbf{I}\sigma_{(\alpha\alpha)}^2)$ , and similarly for additive-by-dominant ( $\alpha\mathbf{d}$ ) and dominant-by-dominant ( $\mathbf{d}\mathbf{d}$ ) effects. Note that under the assumption of HWE, the GBLUP “breeding” model in Vitezica et al. [10] that accounts for additive and dominance effects is a particular case of the model to analyze epistasis in Vitezica et al. [14], which in turn is an extension of the NOIA (natural orthogonal interactions) QTL analysis model by Alvarez-Castro and Carlborg [23]. Note that a “statistical” or “breeding” model implies that the covariance between breeding values and dominance deviations is zero [24]. The NOIA model is orthogonal, which means that genetic effects are defined in such a way that the addition of other genetic effects (e.g. dominance) to a model does not change the definitions of genetic effects (e.g. additive) that were already included in the model [23]. For instance, the

additive effect is always the regression of genotypic value on gene content; the dominance effect is a regression of the remainder on a measure of heterozygosity or identity at the genotype, and so on. The NOIA model of Alvarez-Castro and Carlborg [23] achieves orthogonality automatically due to its construction of the incidence matrices  $\mathbf{M}$  and  $\mathbf{W}$  and their Kronecker products; Vitezica et al. [14] proved that, for analyses at the individual level, these Kronecker products can be reformulated as Hadamard products.

Statistically, orthogonality means that inclusion of new terms in the model does not change estimates of already included effects in an infinitely large population. For instance, in practice, going from an additive to an additive plus dominant model should not change the estimates of additive variance components much. The advantage of using orthogonality in genetics and breeding is that it is the only way to carry out the estimation of breeding values (additive “statistical” effects) in an unambiguous manner i.e. such that they do not depend on other genetic terms that are fitted in the model.

Variance components were estimated for five nested models that added, in succession, additive effects ( $A$ ), dominance effects ( $A + D$ ), additive-by-additive genetic effects ( $A + D + AA$ ), additive-by-dominance genetic effects ( $A + D + AA + AD$ ), and dominance-by-dominance genetic effects ( $A + D + AA + AD + DD$ ). Genetic variances ( $\sigma_A^2$ ,  $\sigma_D^2$ ,  $\sigma_{AA}^2$ ,  $\sigma_{AD}^2$  and  $\sigma_{DD}^2$ ) were estimated by Bayesian methods using Gibbs sampling in the software gibbs2f90 [25], available at <http://nce.ads.uga.edu/wiki/>. In total, 200,000 iterations were run, discarding the first 10,000 and keeping every 100th sample. Convergence was checked by visual inspection of trace plots for the chains. Post-Gibbs analysis included estimation of the effective sample size and the deviance information criteria (DIC) as a goodness-of-fit criterion [26].

### Predictive ability

GBLUP was used to obtain estimated breeding values (EBV) by fixing the variance components that were estimated. Predictive ability of total genetic values for lowly heritable traits is difficult to ascertain, thus for the comparison of models, we proceeded as follows. EBV ( $\hat{\mathbf{g}}_A^w$ ) were obtained from the “whole” dataset (13,369 records from 3619 sows), which included litters from 2000 to 2014. EBV ( $\hat{\mathbf{g}}_A^p$ ) were also computed for a “partial” dataset that included only litters prior to 2010 (10,002 records from 2440 sows). No sow farrowed both before and after 2010 (such sows were excluded from the data set). The three models:  $A$ ,  $A + D$ , and  $A + D + AA + AD + DD$  were used to estimate  $\hat{\mathbf{g}}_A^w$  and  $\hat{\mathbf{g}}_A^p$ . The models were compared using the following three statistics of cross-validation based on method R approaches [27, 28] to detect

bias in subsequent genetic evaluations, and as more recently suggested by [29]:

$$b_0 = (1'\hat{\mathbf{g}}_A^w - 1'\hat{\mathbf{g}}_A^p)/n,$$

$$b_1 = \frac{Cov(\hat{\mathbf{g}}_A^w, \hat{\mathbf{g}}_A^p)}{Var(\hat{\mathbf{g}}_A^p)},$$

$$\rho = \frac{Cov(\hat{\mathbf{g}}_A^w, \hat{\mathbf{g}}_A^p)}{\sqrt{Var(\hat{\mathbf{g}}_A^w) Var(\hat{\mathbf{g}}_A^p)}},$$

where  $b_0 = 0$  is referred to as the bias and should be equal to 0 under unbiasedness;  $b_1$  is a measure of dispersion of EBV, obtained as the slope of the regression of EBV obtained with the whole data on EBV estimated with the partial data, and should be 1; and  $\rho$  measures the relative increase of accuracies of EBV from partial data to whole data and was computed as the correlation of partial data on whole (among two models with different values of  $\rho$ , the model with the highest  $\rho$  was chosen). EBV of only “new” sows that farrowed after 2010 were included in the statistics. These sows have no own or progeny phenotypic information in the partial data and thus, they can be seen as selection candidates at birth.

For comparison purposes, we also investigated the predicted ability of phenotypes of “new” sows for the three models as the correlation  $cor(y^*, \hat{y})$  [30] where  $y^*$  is the corrected phenotypic observation obtained from the “whole” dataset ( $y^* = y - X\hat{\beta} - f\hat{b}$ ), and  $\hat{y}$  is the predicted corrected observation from the “partial” dataset, which is equal to the sum of the estimated genetic values ( $\hat{g}$ ) and the estimated permanent environmental effect ( $\hat{pe}$ ). The estimated genetic value included the estimated additive genetic effects ( $A$  model), the sum of estimated additive and dominant genetic effects ( $\hat{g} = \hat{g}_A + \hat{g}_D$ ) in the  $A + D$  model, and the sum of all genetic effects ( $\hat{g} = \hat{g}_A + \hat{g}_D + \hat{g}_{AA} + \hat{g}_{AD} + \hat{g}_{DD}$ ), in the  $A + D + AA + AD + DD$  model. The regression coefficients of  $y^*$  on  $\hat{y}$  were also estimated.

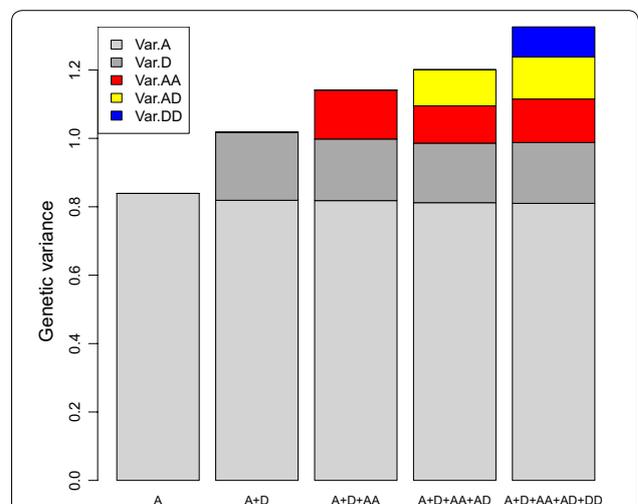
### Results and discussion

Estimates of additive and dominance genetic variances for litter size ranged from  $0.81 \pm 0.12$  to  $0.84 \pm 0.12$ , and from  $0.17 \pm 0.11$  to  $0.20 \pm 0.11$ , respectively for all models ( $A$ ,  $A + D$ ,  $A + D + AA$ ,  $A + D + AA + AD$  and  $A + D + AA + AD + DD$ ). Variance component estimates did not differ among the models (Fig. 1), which empirically illustrates the orthogonality in the partition of the total genetic variance, a property that holds under HWE (which holds in this dataset) and under linkage equilibrium (which holds approximately). Under orthogonality, allele substitution effects contribute to the additive variance, dominance deviations contribute to the dominance variance, etc. and there is

no covariance between the genetic effects. If the model used is not orthogonal (e.g. “genotypic” model in Su et al. [9]), estimates of additive genetic variance may be biased downward when the model is expanded from additive to include dominance and epistatic effects (e.g. Su et al. [9] and Muñoz et al. [31]), and many others). Not using orthogonal models leads to views that are too optimistic on the role of dominance on breeding.

Estimates of epistatic variance ranged from  $0.11 \pm 0.11$  to  $0.14 \pm 0.12$  for the additive-by-additive component, from  $0.11 \pm 0.09$  to  $0.12 \pm 0.09$  for the additive-by-dominant component, and were equal to  $0.09 \pm 0.09$  for the dominant-by-dominant component. As explained in many previous papers [1, 7], the magnitude of the epistatic variances is trivial compared to that of the additive variance. However, non-orthogonal models can result in exaggerated estimates of epistatic variances (e.g. Su et al. [9] and Muñoz et al. [31]). Estimates of epistatic variances had a large standard error in all models ( $A + D + AA$ ,  $A + D + AA + AD$  and  $A + D + AA + AD + AA$ ), which illustrates the difficulties in obtaining good estimates of epistatic variances also from genomic information, even when there are only two-way interactions.

Estimates of narrow-sense heritability for the total number of piglets born per litter were similar between models (Table 1), close to 0.09, and consistent with estimates reported by Varona et al. [32], Nielsen et al. [17] and Guo et al. [18]. Dominance variance expressed as a



**Fig. 1** Estimates of additive (Var.A), dominance (Var.D), additive-by-additive (Var.AA), additive-by-dominance (Var.AD), and dominance-by-dominance (Var.DD) genetic variances for five models that included genomic inbreeding and successively added additive effects ( $A$ ), dominance effects ( $A + D$ ), additive-by-additive effects ( $A + D + AA$ ), additive-by-dominance effects ( $A + D + AA + AD$ ), and dominance-by-dominance effects ( $A + D + AA + AD + DD$ )

proportion of phenotypic variance,  $d^2$ , was about 0.02, as in Misztal et al. [33]. The total epistatic variance (additive-by-additive plus additive-by-dominance plus dominance-by-dominance) expressed as a proportion of phenotypic variance,  $i^2$ , was 0.04. The latter estimates need to be considered with caution due to the low precision of the estimates of epistatic effects. The estimated broad-sense heritability for litter size ( $h^2 + d^2 + i^2$ ) was 0.15, i.e. almost twice the narrow-sense heritability.

The joint posterior distribution of heritability and  $d^2$  or  $i^2$  (Fig. 2) shows that there is no dependency between variance component estimates and thus, the partition of additive and non-additive effects was empirically orthogonal.

Inclusion of non-additive (dominance and epistasis) effects in the model did not have a large effect on estimates of residual variance, but they reduced the

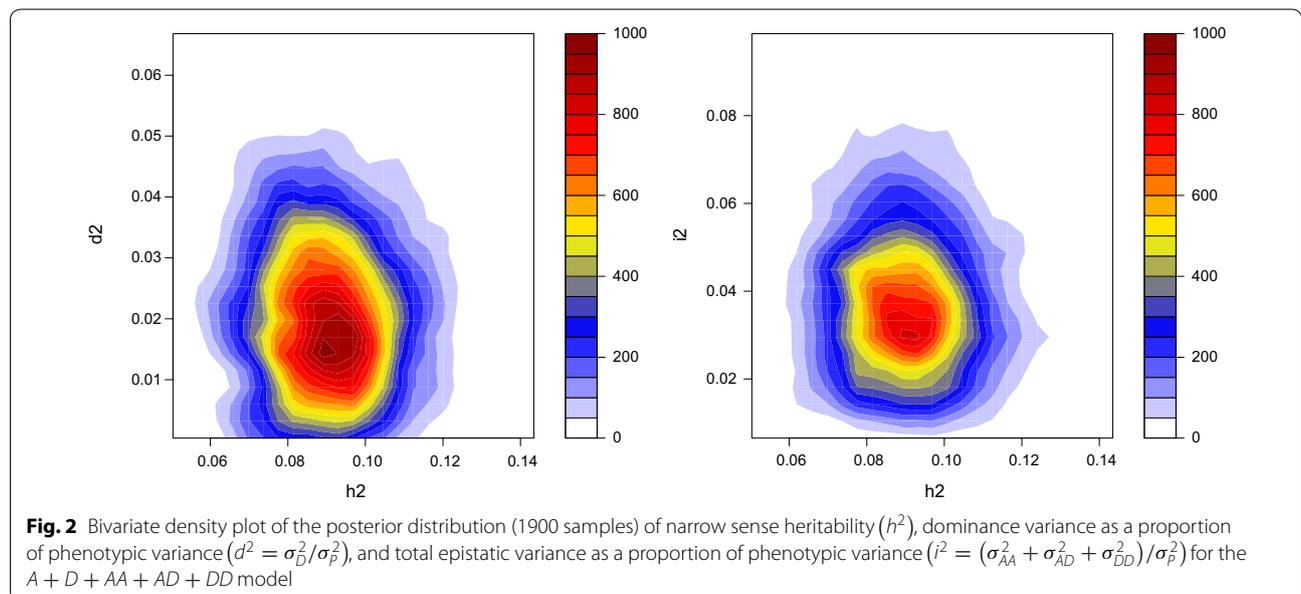
permanent environmental variance (Table 1), which shows that non-additive genetic effects are captured by the permanent environmental effects if they are not included explicitly in the model. Similar results were observed by Aliloo et al. [4] when dominance was included in an additive model.

Only second-order epistatic effects (e.g. additive-by-additive) were included in this study. Although it is tempting to fit high-order epistasis terms given the relative ease of computing the Hadamar products of relationship matrices, caution is needed. First, the products  $\mathbf{G} \odot \mathbf{G} \odot \mathbf{G} \dots$  quickly tend to the identity matrix, in which case there is no hope of distinguishing genetic components from residual effects. Second, partitioning genetic variance into additive and non-additive statistical components does not indicate the importance of additive versus non-additive gene actions [2].

**Table 1 Estimates (and posterior standard deviation) of narrow sense heritability and of variance components for models that included genomic inbreeding and successively added additive effects (A), dominance effects (A + D), additive-by-additive effects (A + D + AA), additive-by-dominance effects (A + D + AA + AD), and dominance-by-dominance effects (A + D + AA + AD + DD)**

Model	$h^2$	$d^{2*}$	$i^{2**}$	$\sigma_{pe}^2$	$\sigma_e^2$
A	0.095 (0.013)			0.931 (0.110)	7.049 (0.116)
A + D	0.093 (0.013)	0.022 (0.013)		0.755 (0.143)	7.053 (0.117)
A + D + AA	0.093 (0.013)	0.020 (0.010)	0.016 (0.015)	0.633 (0.178)	7.052 (0.117)
A + D + AA + AD	0.092 (0.013)	0.020 (0.011)	0.024 (0.014)	0.572 (0.179)	7.051 (0.118)
A + D + AA + AD + DD	0.092 (0.013)	0.019 (0.012)	0.038 (0.017)	0.450 (0.184)	7.054 (0.117)

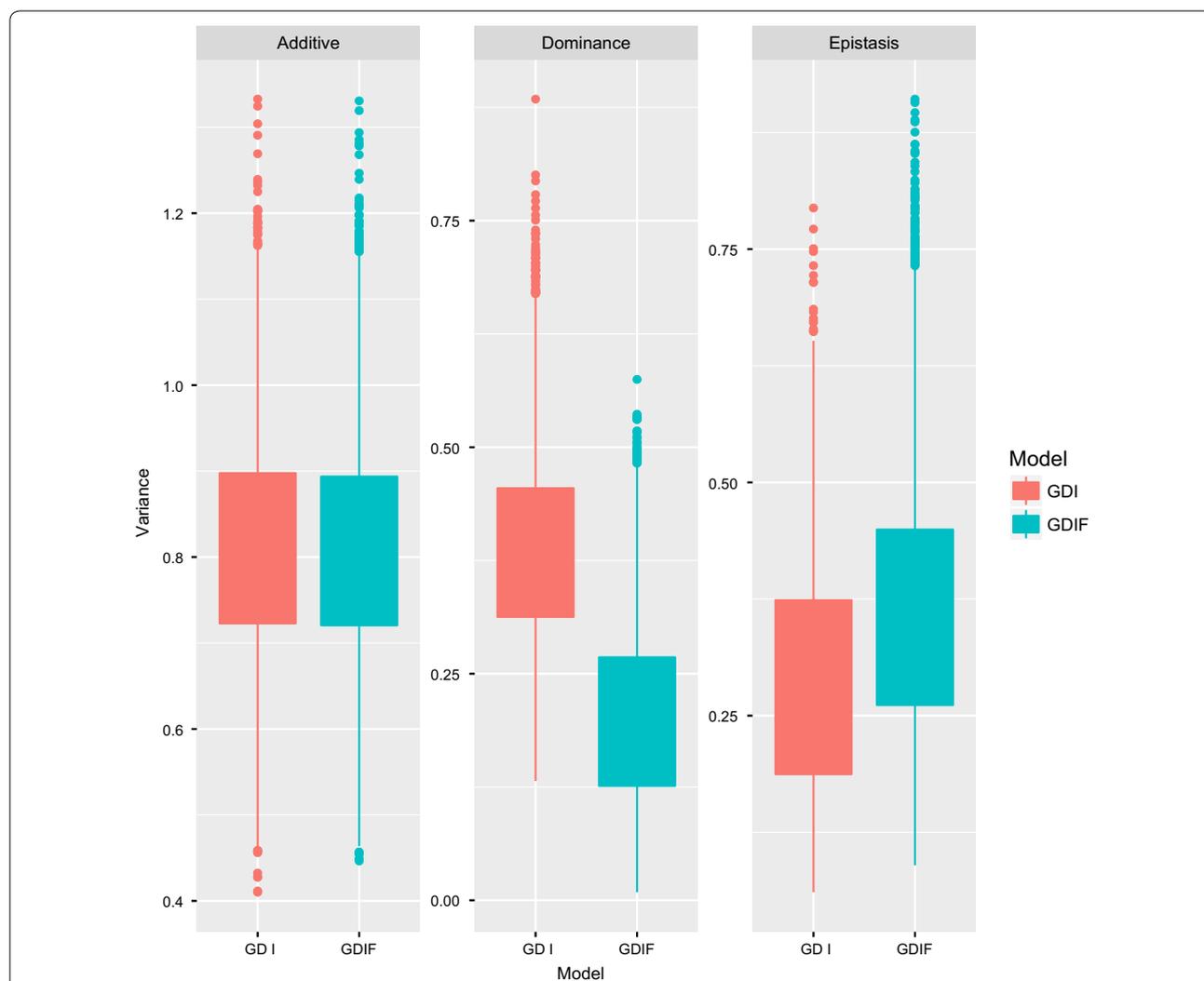
\*  $d^2 = \sigma_D^2 / \sigma_P^2$ , \*\*  $i^2 = \sigma_{AA}^2 / \sigma_P^2$  for the A + D + AA model,  $i^2 = (\sigma_{AA}^2 + \sigma_{AD}^2) / \sigma_P^2$  for the A + D + AA + AD model, and  $i^2 = (\sigma_{AA}^2 + \sigma_{AD}^2 + \sigma_{DD}^2) / \sigma_P^2$  for the A + D + AA + AD + DD model



The estimate of inbreeding depression, based on the inbreeding coefficient calculated as the proportion of homozygosity, was equal to  $-12.33 \pm 2.29$  piglets born per litter for this population. Inbreeding depression expressed as a change in phenotypic mean per 10% increase in inbreeding was equal to  $-1.23$  piglets born. This result shows the importance of inbreeding depression in fitness traits such as total number of piglets born. Estimates of non-additive variance components when inbreeding depression is not fitted in the model are rarely reported but, in our analyses, yielded an upward biased estimate of the dominance variance (i.e. GDI model in Fig. 3). The estimate of dominance variance increased from 0.18 in the GDIF model (including inbreeding) to 0.38 in the GDI model. The estimate of epistatic variance

was slightly affected by inclusion of inbreeding depression in the model. Estimates of additive variance were not affected. Similar results for dominance variance were obtained by Xiang et al. [15] and Aliloo et al. [4]. To take directional dominance into account, inbreeding depression must be included in genetic evaluation models, which has long been known for pedigree analysis (e.g. DeBoer and Hoeschele [34]).

The global fit of the models was studied using DIC. Models with smaller DIC exhibit a better fit after accounting for model complexity. Differences in DIC between models of less than 7 units are considered as irrelevant [35]. The values of DIC were 68,365, 68,370, 68,367, 68,370 and 68,375 for the  $A$ ,  $A + D$ ,  $A + D + AA$ ,  $A + D + AA + AD$  and



**Fig. 3** Estimates (boxplots of posterior distributions) of additive, dominance, and epistatic genetic variances for the  $A + D + AA + AD + DD$  model, with (GDIF) or without (GDI) including genomic inbreeding. The  $A + D + AA + AD + DD$  model involves additive, dominance, additive-by-additive, additive-by-dominance, and dominance-by-dominance effects

$A + D + AA + AD + AA$  models, respectively, i.e. they were similar across models, and models that included dominance and epistasis do not appear to fit the data better than the simplest model.

The predictive ability of the models for selection candidates was assessed using two approaches: the three statistics based on method R and by conventional cross-validation. The method R statistics of bias ( $b_0$ ), slope ( $b_1$ ), and correlation ( $\rho$ ) were based on comparison of EBV obtained from the whole data (up to 2014) and model A with EBV obtained from the partial data and models A,  $A + D$  and  $A + D + AA + AD + AA$ . Similar results (not shown here) were obtained when using the  $A + D$  and  $A + D + AA + AD + DD$  models in the whole dataset. For the correct models,  $b_0 = 0$  and  $b_1 = 1$  are expected. The estimate of  $b_0$  was equal to  $0.019\sigma_A$  across the models and the estimate of  $b_1$  was equal to 1.01, 1.04 and 1.09 in models A,  $A + D$  and  $A + D + AA + AD + DD$ , respectively. These estimates near zero for  $b_0$  and near to 1 for  $b_1$ , suggest that the model is empirically unbiased. The statistic  $\rho$  (accuracy), measured as the correlation between the EBV of selection candidates based on “whole” and “partial” data, was around 0.73 for all models.

Based on conventional cross-validation (Table 2), models A,  $A + D$  and  $A + D + AA + AD + DD$  showed similar predictive ability of phenotypes. Estimates of the regression coefficient of  $y^*$  on  $\hat{y}$  were close to 1 for all three models, which all achieved very similar levels of inflation of total genetic values. No differences in the accuracy of predicting future records for young animals were observed between the three models.

## Conclusions

Using orthogonal relationship matrices, empirically orthogonal estimates of additive, dominance and epistatic variances were obtained for litter size in a pig dataset. The broad-sense heritability for litter size was almost twice the narrow-sense heritability. Genomic models that include non-additive effects must consider simultaneously inbreeding depression based on

**Table 2 Predictive ability of phenotype for different models**

Model	G*	GD**	GDI***
$cor(y^*, \hat{y})^a$	0.135	0.136	0.136
Regression coefficient <sup>b</sup>	0.896	0.922	0.978

<sup>a</sup> Predictive ability ( $cor(y^*, \hat{y})$ ) is measured as the correlation between the corrected phenotypic observation ( $y^*$ ) and the predicted corrected observation ( $\hat{y}$ )

<sup>b</sup> Regression coefficients of  $y^*$  on  $\hat{y}$

\* G is for the additive model, \*\*GD is for the  $A + D$  model, \*\*\* GDI is for the  $A + D + AA + AD + DD$  model

inbreeding in order to obtain unbiased estimates of variance components. Inclusion of epistasis or dominance did not improve the accuracy of prediction of breeding or genotypic values.

## Authors' contributions

ZV and AL proposed the models and methods of analysis. WH provided the pig data. ZV conducted the statistical analyses and wrote the first draft of the manuscript. AR and WH discussed the results, made suggestions and corrections. All authors read and approved the final manuscript.

## Author details

<sup>1</sup> INP ENSAT, UMR 1388 GenPhySE, 31326 Castanet-Tolosan, France. <sup>2</sup> CSIRO Agriculture and Food, 306 Carmody Rd., St Lucia, QLD 4067, Australia. <sup>3</sup> PIC North America, Hendersonville, TN 37075, USA. <sup>4</sup> INRA, UMR 1388 GenPhySE, 31326 Castanet-Tolosan, France.

## Acknowledgements

The authors thank INRA and CSIRO people for hosting their respective visits and for the friendly environment. They also thank Genus plc. (Hendersonville, TN, USA) for the data.

## Competing interests

The authors declare that they have no competing interests.

## Availability of data and materials

The dataset used and analyzed in this study is available from the corresponding author on reasonable request and with the approval of Genus plc.

## Consent for publication

Not applicable.

## Ethics approval and consent to participate

Animal Care and Use Committee approval was not necessary for this study because the data were obtained from an existing database.

## Funding

This work was supported by the INRA SELGEN metaprogram (projects EpiSel and OptiMaGics) and an INRA-CSIRO linkage action. The project was partly supported by the Toulouse Midi-Pyrénées bioinformatics platform.

## Publisher's Note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Received: 19 April 2018 Accepted: 4 December 2018

Published online: 22 December 2018

## References

- Mäki-Tanila A, Hill WG. Influence of gene interaction on complex trait variation with multilocus models. *Genetics*. 2014;198:355–67.
- Huang W, Mackay TFC. The genetic architecture of quantitative traits cannot be inferred from variance component analysis. *PLoS Genet*. 2016;12:e1006421.
- Aliloo H, Pryce JE, González-Recio O, Cocks BG, Hayes BJ. Accounting for dominance to improve genomic evaluations of dairy cows for fertility and milk production traits. *Genet Sel Evol*. 2016;48:8.
- Aliloo H, Pryce JE, González-Recio O, Cocks BG, Goddard ME, Hayes BJ. Including nonadditive genetic effects in mating programs to maximize dairy farm profitability. *J Dairy Sci*. 2017;100:1203–22.
- Toro MA, Varona L. A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol*. 2010;42:33.
- Joshi R, Woolliams JA, Meuwissen THE, Gjøen HM. Maternal, dominance and additive genetic effects in Nile tilapia; influence on growth, fillet yield and body size traits. *Heredity (Edinb)*. 2018;120:452–62.
- Hill WG, Goddard ME, Visscher PM. Data and theory point to mainly additive genetic variance for complex traits. *PLoS Genet*. 2008;4:e1000008.

8. Hill WG. "Conversion" of epistatic into additive genetic variance in finite populations and possible impact on long-term selection response. *J Anim Breed Genet.* 2017;134:196–201.
9. Su G, Christensen OF, Ostersen T, Henryon M, Lund MS. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One.* 2012;7:e45293.
10. Vitezica ZG, Varona L, Legarra A. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics.* 2013;195:1223–30.
11. Nishio M, Satoh M. Including dominance effects in the genomic BLUP method for genomic evaluation. *PLoS One.* 2014;9:e85792.
12. Jiang Y, Reif JC. Modeling epistasis in genomic selection. *Genetics.* 2015;201:759–68.
13. Lynch M, Walsh B. *Genetics and analysis of quantitative traits.* New York: Sinauer associates Inc.; 1998.
14. Vitezica ZG, Legarra A, Toro MA, Varona L. Orthogonal estimates of variances for additive, dominance, and epistatic effects in populations. *Genetics.* 2017;206:1297–307.
15. Xiang T, Christensen OF, Vitezica ZG, Legarra A. Genomic evaluation by including dominance effects and inbreeding depression for purebred and crossbred performance with an application in pigs. *Genet Sel Evol.* 2016;48:92.
16. Silió L, Rodríguez MC, Fernández A, Barragán C, Benítez R, Óvilo C, et al. Measuring inbreeding and inbreeding depression on pig growth from pedigree or SNP-derived metrics. *J Anim Breed Genet.* 2013;130:349–60.
17. Nielsen B, Su G, Lund MS, Madsen P. Selection for increased number of piglets at d 5 after farrowing has increased litter size and reduced piglet mortality. *J Anim Sci.* 2013;91:2575–82.
18. Guo X, Christensen OF, Ostersen T, Wang Y, Lund MS, Su G. Improving genetic evaluation of litter size and piglet mortality for both genotyped and nongenotyped individuals using a single-step method. *J Anim Sci.* 2015;93:503–12.
19. Bidanel JP. Biology and genetics of reproduction. In: Rothschild MF, Ruvinsky A, editors. *The genetics of the pig.* 2nd ed. London: AB International; 1998. p. 313–43.
20. Aguilar I, Misztal I, Tsuruta S, Legarra A, Wang H. PREGSF90–POSTGSF90: computational tools for the implementation of single-step genomic selection and genome-wide association with ungenotyped individuals in BLUPF90 programs. In: *Proceedings of the 10th world congress of genetics applied to livestock production: 17–22 Aug 2014, Vancouver.*
21. Hill WG, Mäki-Tanila A. Expected influence of linkage disequilibrium on genetic variance caused by dominance and epistasis on quantitative traits. *J Anim Breed Genet.* 2015;132:176–86.
22. VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci.* 2008;91:4414–23.
23. Alvarez-Castro JM, Carlborg O. A unified model for functional and statistical epistasis and its application in quantitative trait loci analysis. *Genetics.* 2007;176:1151–67.
24. Falconer DS, Mackay TFC. *Introduction to quantitative genetics.* New York: Longman; 1996.
25. Misztal I, Tsuruta S, Strabel T, Auvray B, Druet T, Lee DH. BLUPF90 and related programs (BGF90). In: *Proceedings of the 7th world congress of genetics applied to livestock production, 19–23 Aug 2002, Montpellier.* Communication No 28-07.
26. Spiegelhalter DJ, Best NG, Carlin BP, Van Der Linde A. Bayesian measures of model complexity and fit. *J R Stat Soc Ser B Stat Methodol.* 2002;64:583–639.
27. Reverter A, Golden BL, Bourdon RM, Brinks JS. Method R variance components procedure: application on the simple breeding value model. *J Anim Sci.* 1994;72:2247–53.
28. Reverter A, Golden BL, Bourdon RM, Brinks JS. Technical note: detection of bias in genetic predictions. *J Anim Sci.* 1994;72:34–7.
29. Legarra A, Reverter A. Can we frame and understand cross-validation results in animal breeding? In: *Proceedings of the 22nd conference of the association for the advancement of animal breeding and genetics, 2–5 July 2017, Townsville.* p. 2273–80.
30. Legarra A, Robert-Granié C, Manfredi E, Elsen JM. Performance of genomic selection in mice. *Genetics.* 2008;180:611–8.
31. Muñoz PR, Resende MFR, Gezan SA, Resende MDV, de los Campos G, Kirst M, et al. Unraveling additive from nonadditive effects using genomic relationship matrices. *Genetics.* 2014;198:1759–68.
32. Varona L, Sorensen D, Thompson R. Analysis of litter size and average litter weight in pigs using a recursive model. *Genetics.* 2007;177:1791–9.
33. Misztal I, Varona L, Culbertson M, Bertrand JK, Mabry J, Lawlor TJ, et al. Studies on the value of incorporating the effect of dominance in genetic evaluations of dairy cattle, beef cattle and swine. *Biotechnol Agron Soc Environ.* 1998;2:227–33.
34. De Boer I, Hoeschele I. Genetic evaluation methods for populations with dominance and inbreeding. *Theor Appl Genet.* 1993;86:245–58.
35. Plummer M, Best N, Cowles K, Vines K. CODA: convergence diagnosis and output analysis for MCMC. *R News.* 2006;6:7–11.

Ready to submit your research? Choose BMC and benefit from:

- fast, convenient online submission
- thorough peer review by experienced researchers in your field
- rapid publication on acceptance
- support for research data, including large and complex data types
- gold Open Access which fosters wider collaboration and increased citations
- maximum visibility for your research: over 100M website views per year

At BMC, research is always in progress.

Learn more [biomedcentral.com/submissions](https://biomedcentral.com/submissions)

