

# Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers

Guosheng Su<sup>1\*</sup>, Ole F. Christensen<sup>1</sup>, Tage Ostensen<sup>2</sup>, Mark Henryon<sup>2,3</sup>, Mogens S. Lund<sup>1</sup>

**1** Department of Molecular Biology and Genetics, Aarhus University, AU-Foulum, Tjele, Denmark, **2** Pig Research Centre, Danish Agricultural & Food Council, Copenhagen, Denmark, **3** School of Animal Biology, University of Western Australia, Crawley, Western Australia, Australia

## Abstract

Non-additive genetic variation is usually ignored when genome-wide markers are used to study the genetic architecture and genomic prediction of complex traits in human, wild life, model organisms or farm animals. However, non-additive genetic effects may have an important contribution to total genetic variation of complex traits. This study presented a genomic BLUP model including additive and non-additive genetic effects, in which additive and non-additive genetic relation matrices were constructed from information of genome-wide dense single nucleotide polymorphism (SNP) markers. In addition, this study for the first time proposed a method to construct dominance relationship matrix using SNP markers and demonstrated it in detail. The proposed model was implemented to investigate the amounts of additive genetic, dominance and epistatic variations, and assessed the accuracy and unbiasedness of genomic predictions for daily gain in pigs. In the analysis of daily gain, four linear models were used: 1) a simple additive genetic model (MA), 2) a model including both additive and additive by additive epistatic genetic effects (MAE), 3) a model including both additive and dominance genetic effects (MAD), and 4) a full model including all three genetic components (MAED). Estimates of narrow-sense heritability were 0.397, 0.373, 0.379 and 0.357 for models MA, MAE, MAD and MAED, respectively. Estimated dominance variance and additive by additive epistatic variance accounted for 5.6% and 9.5% of the total phenotypic variance, respectively. Based on model MAED, the estimate of broad-sense heritability was 0.506. Reliabilities of genomic predicted breeding values for the animals without performance records were 28.5%, 28.8%, 29.2% and 29.5% for models MA, MAE, MAD and MAED, respectively. In addition, models including non-additive genetic effects improved unbiasedness of genomic predictions.

**Citation:** Su G, Christensen OF, Ostensen T, Henryon M, Lund MS (2012) Estimating Additive and Non-Additive Genetic Variances and Predicting Genetic Merits Using Genome-Wide Dense Single Nucleotide Polymorphism Markers. *PLoS ONE* 7(9): e45293. doi:10.1371/journal.pone.0045293

**Editor:** Abraham A. Palmer, University of Chicago, United States of America

**Received:** June 13, 2012; **Accepted:** August 14, 2012; **Published:** September 13, 2012

**Copyright:** © 2012 Su et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

**Funding:** The work was performed in the project (grant no. 3405-11-0279) funded by the Danish Ministry of Food, Agriculture and Fisheries, Pig Research Centre and Aarhus University. The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

**Competing Interests:** The authors have declared that no competing interests exist.

\* E-mail: guosheng.su@agrsci.dk

## Introduction

Non-additive genetic variation results from interactions between genes. Interactions between genes at the same locus are called dominance, and interactions between genes at different loci are called epistasis. Although many studies have shown that non-additive effects have a substantial contribution to variation of complex traits [1–3], this source of variation is generally ignored in the genetic evaluation of complex traits.

Genome-wide dense single nucleotide polymorphism (SNP) markers have been widely used for association analysis [4–8] and genomic selection [9–12]. Unlike association analysis which aims at identifying quantitative trait loci (QTL) or chromosome regions with significant effect on the trait of interest, genomic selection focuses on predicting breeding values (total additive genetic effects). Similarly, when considering non-additive genetic effects, association analysis tries to find interactions among the specific genes that have a large effect on the trait of interest, while genomic selection pays attention to total non-additive genetic variations.

Many statistical models and algorithms have been proposed to predict breeding values using genome-wide dense markers, which

differ in the assumption of distributions of SNP effects. Best linear unbiased prediction (BLUP) models [13–15] assume that effects of all SNP are normally distributed with equal variance. Variable selection models [12,14,16,17] assume that marker effects have a thick-tailed distribution or a mixture distribution. Simulation studies with assumption that few QTL affect the trait of interest have shown that variable selection models are superior over BLUP models [14,18,19]. However, studies based on real dairy cattle and pig data indicate that BLUP models performed as well as variable selection models for most traits [11,20,21]. Therefore BLUP models have become popular approaches in practical genomic evaluations because they are simple and have low computational demands.

There are two BLUP models which have been widely used for genomic predictions. One estimates marker effects using random regression on marker genotypes and genomic breeding value of an individual is calculated as the sum of estimated marker effects (hereafter called as SNP-BLUP). The other estimates genomic breeding value directly using a marker-based relationship matrix (hereafter denoted as GBLUP). It has been shown that the GBLUP model is equivalent to the SNP-BLUP model [22,23].

One of the advantages of the GBLUP model is that the model can use the framework of traditional pedigree-based BLUP models and is easy to use the information of non-genotyped animals through a combined relationship matrix [24–27].

Similar to traditional genetic evaluations, genomic predictions are usually carried out using a model that ignores non-additive effects. The hypothesis of this study is that statistical models which include additive and non-additive genetic effects will predict genetic merit more accurately and with less bias, when non-additive genetic effects have a substantial contribution to the genetic variation.

The objectives of this study were twofold. The first was to describe an approach to estimate additive and non-additive genetic variations and predict genetic values for complex traits using models integrating additive and non-additive genomic relationship matrices. The second was to estimate additive, epistatic and dominance genetic variances and assess the accuracy of genomic predictions for daily gain in Danish Duroc pigs using models with or without including non-additive genetic effects.

## Materials and Methods

### Genomic BLUP model for additive and non-additive genetic effects

A linear mixed model including additive and non-additive genetic effects can be written as:

$$\mathbf{y} = \mathbf{X}\mathbf{b} + \mathbf{Z}_a\mathbf{a} + \mathbf{Z}_i\mathbf{i} + \mathbf{Z}_d\mathbf{d} + \mathbf{e}$$

where  $\mathbf{y}$  is the vector of observations,  $\mathbf{b}$  is the vector of non-genetic effects,  $\mathbf{a}$  is the vector of additive genetic effects,  $\mathbf{i}$  is the vector of epistatic effects (second order epistasis in this study),  $\mathbf{d}$  is the vector of dominance effects, and  $\mathbf{e}$  is the vector of random residuals.

It is assumed that

$$\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2), \mathbf{i} \sim N(0, \mathbf{G}_{aa}\sigma_{aa}^2), \mathbf{d} \sim N(0, \mathbf{D}\sigma_d^2), \text{ and } \mathbf{e} \sim N(0, \mathbf{I}\sigma_e^2),$$

where  $\sigma_a^2$  is the additive genetic variance,  $\sigma_{aa}^2$  is the epistatic variance,  $\sigma_d^2$  is the dominance variance,  $\sigma_e^2$  is the residual variance,  $\mathbf{I}$  is an identity matrix,  $\mathbf{G}$ ,  $\mathbf{G}_{aa}$  and  $\mathbf{D}$  are the additive, epistatic and dominance genetic relationship matrices, respectively. These matrices can be constructed from either the information of pedigree or the information of genome-wide markers. The marker-based relationship matrices have the advantage to capture both the Mendelian segregation and the genetic links through unknown common ancestors, which are not available in the known pedigree. The present study will demonstrate the calculation of additive, dominance and epistatic genetic relationship matrices based on genome-wide markers. In the context, the three marker-based relationship matrices will be denoted as additive, epistatic and dominance genomic relationship matrices.

**Additive genomic relationship matrix.** The additive genomic relationship matrix  $\mathbf{G}$  can be constructed using SNP marker information according to the previous studies [13,15].

Briefly,  $\mathbf{G} = \frac{\mathbf{M}\mathbf{M}'}{\sum 2p_iq_i}$ , where  $\mathbf{M}$  is a  $n \times m$  matrix ( $n$  = number of animals,  $m$  = number of marker loci) which specifies SNP genotype coefficients at each locus. The coefficients of the  $i^{\text{th}}$  column in the  $\mathbf{M}$  matrix are  $(0 - 2p_i)$  for genotype  $A_1A_1$ ,  $(1-2p_i)$  for  $A_1A_2$ , and  $(2-2p_i)$  for  $A_2A_2$ , where  $q_i$  and  $p_i$  are the frequencies of allele 1 ( $A_1$ ) and allele 2 ( $A_2$ ) at locus  $i$ , respectively.

**Epistatic genomic relationship matrix.** According to the previous study [28], the epistatic genomic relationship matrix can be derived from additive genomic relationship matrix. When

considering only second order epistasis (i.e., additive by additive interactions) and ignoring inbreeding, the epistatic genomic relationship matrix is:  $\mathbf{G}_{aa} \approx \mathbf{G}\#\mathbf{G}$ , where  $\#$  denotes the Hadamard product operation.

**Dominance genomic relationship matrix.** The dominance genomic relationship matrix  $\mathbf{D}$  can be derived as follows. Denote  $u$  as the dominance value at a single locus and  $h_0$  as the heterozygosity of an individual at the locus. The dominance effect of an individual at a locus can be described as  $\mathbf{d}_0 = h_0\mathbf{u}$ . For a locus with two alleles (say  $A_1$  and  $A_2$ ),  $h_0 = 0$  for the homozygous genotypes  $A_1A_1$  and  $A_2A_2$ , and  $h_0 = 1$  for the heterozygous genotype  $A_1A_2$ . Since  $h_0$  is either 0 or 1,  $h_0 \sim \text{Bernoulli}(1, 2pq)$ . Therefore,  $E(h_0) = 2pq$  and  $\sigma_{h_0}^2 = 2pq(1 - 2pq)$ .

To simplify the algorithm in the above linear mixed models, it is assumed that dominance effects ( $\mathbf{d}$ ) are normally distributed with mean equal to zero. In order to meet this assumption, dominance effects are expressed as deviations from the population mean,  $\mathbf{h} = (h_0 - 2pq)$ . Thus, let  $\mathbf{d} = \mathbf{h}\mathbf{u}$ , then  $E(\mathbf{h}) = 0$ ,  $E(\mathbf{d}) = 0$ ,  $\sigma_h^2 = 2pq(1 - 2pq)$  and  $\sigma_d^2 = \sigma_h^2\sigma_u^2$ , where  $\sigma_u^2$  is the variance of dominance values.

Expanding to  $m$  loci, the genome-wide dominance effect of an individual is  $\mathbf{d} = \mathbf{h}\mathbf{u}$ , where  $\mathbf{h}$  is the  $m$ -dimensional vector of heterozygosity coefficients, and  $\mathbf{u}$  is the  $m$ -dimensional vector of dominance values. Assuming that dominance values at different loci are identically and independently distributed normal variables, the variance of genome-wide dominance effects in a Hardy-Weinberg equilibrium population is

$$\sigma_d^2 = \mathbf{V}(\mathbf{h}\mathbf{u}) = \sum_{i=1}^m 2p_iq_i(1 - 2p_iq_i)\sigma_u^2.$$

Expanding to  $n$  individuals and  $m$  loci,

$$\mathbf{d} = \mathbf{H}\mathbf{u},$$

where  $\mathbf{d}$  is the  $n$ -dimensional vector of dominance effects,  $\mathbf{H}$  is the  $n \times m$  matrix of heterozygosity coefficients with element  $h_{ki} = 0 - 2p_iq_i$  if individual  $k$  is homozygous, and  $h_{ki} = 1 - 2p_iq_i$  if individual  $k$  is heterozygous at locus  $i$ . The covariance structure of  $\mathbf{d}$  is

$$\mathbf{V}(\mathbf{d}) = \mathbf{H}\mathbf{H}'\sigma_u^2 = \frac{\mathbf{H}\mathbf{H}'}{\sum 2p_iq_i(1 - 2p_iq_i)}\sigma_d^2.$$

Consequently, the dominance genomic relationship matrix  $\mathbf{D}$  is,

$$\mathbf{D} = \frac{\mathbf{H}\mathbf{H}'}{\sum 2p_iq_i(1 - 2p_iq_i)}.$$

By centering  $h_{ki}$  with  $2p_iq_i$  and scaling  $\mathbf{H}\mathbf{H}'$  with  $\sum 2p_iq_i(1 - 2p_iq_i)$ , the  $\mathbf{D}$  matrix has the properties that the expectation of an off-diagonal element is zero for two unrelated individuals, and the expectation of a diagonal element is one for a non-inbred individual. The  $\mathbf{D}$  matrix is a realized dominance relationship matrix and is analogous to the pedigree-based numerator dominance relationship matrix.

**Marker data and phenotypic data of daily gain in pigs**

**Marker data.** SNP marker data were obtained from 1911 Danish Duroc pigs (most were boars) which were genotyped using Illumina PorcineSNP60 BeadChip (Illumina, San Diego, CA). The SNP data were edited using the following criteria: 1) Minor allele frequency higher than 5%; 2) locus call rate larger than 0.95; and 3) individual animal call rate larger than 0.95. In addition, if the GenCall score of a single SNP in an animal was less than 0.65, the SNP in this animal was treated as missing and the corresponding genotype coefficient was replaced by the expected genotype coefficient at this locus. After editing, there were 26,142 SNP markers available for 1,911 pigs (born from 1996–2009, with 77% born during 2006–2009).

**Phenotypic data.** The analyzed trait was average daily gain (DG) from 30 kg to 100 kg. More details about the data can be found in the previous study [21]. Five datasets were used in the current analysis. DG of 339,393 individuals born during 1992–2009 (DATA<sub>all</sub>) were used to calculate corrected phenotypic values of DG (see detail below). The corrected DG of the 1,911 genotyped pigs (DATA<sub>gen</sub>) were used to estimate additive and non-additive genetic variances. The data were further divided into reference dataset (DATA<sub>ref</sub>, n = 1,484) and test dataset (DATA<sub>test</sub>, n = 427) by a cut-off date June 1, 2008 (birth date). DATA<sub>ref</sub> were used to predict breeding value of the genotyped pigs in DATA<sub>test</sub>.

Corrected phenotypic values of daily gains (DG<sub>c</sub>), instead of original observations, were used as response variables to estimate additive and non-additive genetic variances and to predict genetic effects using SNP markers. The reason for using DG<sub>c</sub> as response variables was to reduce noise by removing contemporary group effects which could be estimated much more accurately using a large dataset including all contemporaries and relatives, rather than using only genotyped animals. The contemporary group effects were estimated using a traditional pedigree-based linear model including sex, herd-week-section, pen, litter and additive genetic effects as well as random residuals. The DG<sub>c</sub> was defined as original observations of daily gain adjusted for all non-genetic effects except for litter effect because based on this model it included part of non-additive genetic effects that should not be corrected for.

**Statistical analysis of daily gain**

Four linear mixed models were used to estimate variance components and predict genetic effects, based on DG<sub>c</sub> of genotyped animals.

$$y = 1\mu + W_1l + Z_a a + e \quad (MA)$$

$$y = 1\mu + W_1l + Z_a a + Z_i i + e \quad (MAE)$$

$$y = 1\mu + W_1l + Z_a a + Z_d d + e \quad (MAD)$$

$$y = 1\mu + W_1l + Z_a a + Z_i i + Z_d d + e \quad (MAED)$$

where **y** is the vector of DG<sub>c</sub>, **μ** is the intercept, **l** is the vector of litter effects and is assumed that  $l \sim N(0, I\sigma_l^2)$ , and the definitions of **a**, **i**, **d** and **e** are the same as stated above.

The analyses were carried out applying the average information restricted maximum likelihood algorithm [29] implemented in the DMU package (<http://dmu.agrsci.dk>).

**Model Validation**

Goodness of fit for each model was evaluated by likelihood value based on the dataset DATA<sub>gen</sub>. The superiority of an alternative model over model MA was tested using a likelihood ratio test. The predictive ability of the model (with respect to accuracy and unbiasedness) was evaluated by comparing predictions and DG<sub>c</sub> of animals in the test dataset (DATA<sub>test</sub>). Prediction accuracy was measured as the correlation between predicted genetic values and DG<sub>c</sub>. Both predictions of additive genetic effect (breeding value) and total genetic value (defined as the sum of the genetic effects in the model) were evaluated. Hotelling-Williams t-test [30,31] was implemented to test if the correlations obtained from these prediction methods were significantly different. By definition, the reliability of predicted additive genetic effects is the squared correlation between predicted and true additive genetic effects,  $r_a^2 = \frac{Cov^2(\hat{a}, a)}{\sigma_a^2 \sigma_a^2}$ . Since true additive genetic effects were unknown, the reliability of predicted additive genetic effects was calculated as the squared correlation between predicted additive genetic effects and DG<sub>c</sub>, divided by heritability of DG<sub>c</sub>, i.e.,  $r_a^2 = \frac{Cov^2(\hat{a}, DG_c)}{\sigma_a^2 \sigma_{DG_c}^2 h_{DG_c}^2} = \frac{Cov^2(\hat{a}, a + residual)}{\sigma_a^2 \sigma_a^2} = \frac{Cov^2(\hat{a}, a)}{\sigma_a^2 \sigma_a^2}$ . Here, the heritability  $h_{DG_c}^2$  was the narrow-sense heritability estimated from the full model MAED. Unbiasedness of genomic predictions was measured using the regression of DG<sub>c</sub> on the genomic predictions. A necessary condition for unbiased predictions is that the regression coefficient does not deviate significantly from one.

**Results**

**Mean and standard deviation of daily gain**

The mean and standard deviation of DG<sub>c</sub> (the corrected daily gain) were significantly different in the different datasets (Table 1). This was due to continuous selection for this trait, which led to a continuous genetic progress. Animals with records in the whole dataset were born during the period from 1992 to 2009 (almost the same number of animals per year), while most genotyped animals were born during the period from 2006 to 2009. Thus, the mean DG<sub>c</sub> in the whole dataset was lower than that in the dataset of genotyped animals. Similarly, the animals in the reference dataset were born before those in the test dataset and consequently had a lower mean in DG<sub>c</sub>. The selection also resulted in a larger overall standard deviation of DG<sub>c</sub> in the dataset covering a longer period, because the standard deviation included the variation of year means. Therefore, the within-year standard deviation is a more

**Table 1.** Mean, total standard deviation (SD<sub>t</sub>) and within-year standard deviation (SD<sub>w</sub>) of the corrected phenotypic values of daily gain in different datasets.

Dataset	N	Mean	SD <sub>t</sub>	SD <sub>w</sub>
All	339,393	964	134	74
Genotyped	1,911	1,134	78	67
Reference	1,484	1,125	77	64
Test	427	1,164	74	74

doi:10.1371/journal.pone.0045293.t001

appropriate measure of the variation of this trait. The coefficients of within-year variation for  $DG_c$  in different datasets ranged from 5.7% to 7.7%.

**Estimates of variance components and heritability**

The dominance variance accounted for 5.6% of the phenotypic variance (Table 2). The epistatic variance accounted for about 9.5% of the phenotypic variance, but the estimate had a large standard error, and thus was not statistically significantly different from zero. Model MA resulted in the largest estimates of additive genetic, litter and residual variances, while model MAED led to smallest estimates of these variances, and estimates from model MAE and MAD were intermediate in size. The results suggested that when a model excluded non-additive genetic effects, the variation due to non-additive genetic effects was distributed to other variance components in the model. Estimates of narrow-sense heritability (proportion of additive genetic variance to phenotypic variance) were 0.397, 0.373, 0.379 and 0.357 from model MA, MAE, MAD and MAED, respectively. Based on model MAED, the estimate of broad-sense heritability (proportion of the sum of additive, epistatic and dominance genetic variances to phenotypic variance) was 0.506.

**Goodness of fit**

Measures of goodness of fit are given in Table 3. Model MA had the largest -2 log likelihood, followed by MAE and then by MAD, whereas model MAED had the lowest -2 log likelihood and thus fitted the data best. However, the likelihood ratio test showed that model MAE did not fit data significantly better than MA. On the other hand, model MAD and MAED were superior to model MA at a level approaching to statistical significance ( $P=0.054$  for MAD vs. MA,  $P=0.067$  for MAED vs. MA). These results indicated that the goodness of fit was improved by including dominance effects, but no detectable improvement was observed by including epistatic effects.

**Table 2.** Estimates of additive genetic variance ( $\sigma_a^2$ ), epistatic variance ( $\sigma_{aa}^2$ ), dominance variance ( $\sigma_d^2$ ), litter variance ( $\sigma_l^2$ ), residual variance ( $\sigma_e^2$ ) and their standard errors, and the proportions of these variances ( $h_a^2, h_{aa}^2, h_d^2, l^2$ ) to phenotypic variance (defined as the sum of variance components in the model).

Parameters	MA	MAE	MAD	MAED
$\sigma_a^2$	2,176±241	2,029±255	2,081±246	1,942±260
$\sigma_{aa}^2$	-	529±429	-	506±429
$\sigma_d^2$	-	-	309±175	303±175
$\sigma_l^2$	604±231	523±237	542±231	465±238
$\sigma_e^2$	2,707±241	2,362±371	2,557±252	2,231±376
$h_a^2$	0.397**	0.373**	0.379**	0.357**
$h_{aa}^2$	-	0.098	-	0.093
$h_d^2$	-	-	0.056*	0.056*
$l^2$	0.110*	0.096*	0.099*	0.085*

\*: Significantly differ from 0 at  $P<0.05$ .  
 \*\*: Significantly differ from 0 at  $P<0.01$   
 doi:10.1371/journal.pone.0045293.t002

**Table 3.** -2log likelihood,  $\chi^2$  value and the corresponding P-value of likelihood ratio.

Model	-2logL	$\chi^2$ -value <sup>a)</sup>	P-value
MA	18019.6		
MAE	18017.8	1.8	0.180
MAD	18015.9	3.7	0.054
MAED	18014.2	5.4	0.067

a)  $\chi^2 = -2 \ln \left( \frac{\text{likelihood for MA}}{\text{likelihood for alternative model}} \right)$   
 doi:10.1371/journal.pone.0045293.t003

**Accuracy of prediction**

Model MAED led to the highest correlation between predicted total genetic values and  $DG_c$ , followed by MAD, then followed by MAE, and MA yielded the lowest correlation (Table 4). The same pattern was observed in correlations between predicted additive genetic effects (which is typically defined as breeding value) and  $DG_c$ . Reliabilities of predicted additive genetic effects were 28.5%, 28.8%, 29.2% and 29.5% for models MA, MAE, MAD and MAED, respectively. The Hotelling-Williams t test showed that the difference in accuracies of predictions using these models were statistically significant from zero, except for predictions of total genetic values between models MA and MAE. The results indicated that including non-additive genetic effects in a prediction model improve accuracy of genomic predictions.

**Unbiasedness of prediction**

As shown in Table 5, the range of regression coefficients of  $DG_c$  on predicted total genetic value was between 0.927 (model MA) and 0.985 (model MAED), the regression coefficients of  $DG_c$  on predicted additive genetic values ranged from 0.927 (model MA) to 1.029 (model MAED). For all models, the regression coefficients were not significantly different from 1, indicating the predictions were not significantly biased. However, the regression coefficients for the predictions using the models that included the non-additive genetic effects were closer to 1, suggesting that these models slightly improved the unbiasedness of genomic predictions.

**Discussion**

This study described a novel approach to model additive, epistatic and dominance genetic effect. The approach, by constructing additive and non-additive genetic relationship matri-

**Table 4.** Correlation between corrected daily gain ( $DG_c$ ) and predicted total genetic value (GTV, defined as the sum of genetic effects in the model), between  $DG_c$  and estimated additive genetic effect (GBV), and reliability of GBV ( $R^2_{GBV}$ ), for the animals in the test dataset.

Model	Cor(GTV, $DG_c$ )	Cor(GBV, $DG_c$ )	$R^2_{GBV}$ (%)
MA	0.319 <sup>a</sup>	0.319 <sup>a</sup>	28.5 <sup>a</sup>
MAE	0.320 <sup>a</sup>	0.321 <sup>b</sup>	28.8 <sup>b</sup>
MAD	0.330 <sup>b</sup>	0.323 <sup>c</sup>	29.2 <sup>c</sup>
MAED	0.331 <sup>c</sup>	0.325 <sup>d</sup>	29.5 <sup>d</sup>

a-d) Within a column, estimates without a common superscript differ significantly ( $P<0.05$ ), according to the Hotelling-Williams t test.  
 doi:10.1371/journal.pone.0045293.t004

**Table 5.** Regression coefficients ( $\pm$  standard errors) of corrected daily gain on predicted total genetic value (GTV) and on predicted breeding value (GBV) for the animals in the test dataset.

Model	Reg. on GTV	Reg. on GBV
MA	0.927 $\pm$ 0.134	0.927 $\pm$ 0.134
MAE	0.954 $\pm$ 0.137	0.981 $\pm$ 0.140
MAD	0.959 $\pm$ 0.133	0.983 $\pm$ 0.140
MAED	0.985 $\pm$ 0.136	1.029 $\pm$ 0.147

doi:10.1371/journal.pone.0045293.t005

ces using information of genome-wide SNP markers, is generally applicable for different populations with or without pedigree information. The method was used to investigate the variance components of additive and non-additive genetic effects and the accuracy of genomic predictions for DG in Danish Duroc population. The estimates of dominance variance and epistatic variance indicated that non-additive genetic effects had a substantial contribution to total genetic variation of this trait. The models including non-additive genetic effects predicted breeding values more accurately and unbiasedly, compared with a model ignoring non-additive genetic effects.

#### Additive and non-additive genomic relationship matrices

Genome-wide dense markers provide a new approach to detect non-additive genetic variation and predict genetic merit. Additive genomic relationship matrix [13,15] has been widely used for genomic prediction applying a linear mixed model [20,32,33]. The present study firstly demonstrated the method to construct dominance relationship matrix using genome-wide dense SNP markers. Following the previous study [28], epistatic genomic relationship matrix can be easily derived from additive genomic relationship matrix. The genomic relationship matrices represent the realized rather than expected sharing of ancestral genes. Compared to the pedigree-based relationship matrix, the genomic relationship matrix can capture both the Mendelian segregation and the genetic links through unknown common ancestors, which are not available in the known pedigree. Furthermore the genomic relationship matrices are applicable for different populations with or without pedigree information, which is particularly advantageous in studies on wild populations or human populations [34,35].

The proposed method to construct epistatic genomic relationship matrix is an approximate approach and only second order epistasis (additive by additive interactions) is considered. Following previous report [28], higher order epistatic genomic relationship matrix can be approximately calculated as particular Hadamard product, such as  $\mathbf{G}\#\mathbf{G}\#\mathbf{D}$  for additive by additive by dominance genomic relationship matrix,  $\mathbf{G}\#\mathbf{G}\#\mathbf{G}$  for additive by additive by additive genomic relationship matrix. However, when higher order epistasis is involved in the analysis, more data information is required. A precise epistatic relationship matrix is difficult to be obtained when considering a large number of markers. Epistasis among a few specific markers can be modeled in a desired way [36–38]. However, instead of dealing with a small proportion of non-additive genetic variation from a few specific markers, the current study tries to use a feasible approach to account for overall non-additive genetic variation as accurate as possible.

With additive and non-additive genetic relationship matrices, additive and non-additive genetic variances and genetic values can

be easily estimated using a typical linear mixed model, such as a GBLUP model. A GBLUP model is equivalent to a linear random regression model assuming that effects of all SNP are normally distributed with equal variance. A GBLUP model may be not satisfactory in the situation that few markers have large effect while the most markers have null effect or very small effect. However, experiences with real dairy cattle data and pig data indicate that a BLUP model performed well for most traits. Therefore, the approach proposed in this study could be not an optimal but a simple and feasible approach for estimating additive and non-additive variances and predicting genomic breeding values.

#### Additive and non-additive genetic variances of daily gain

Based on the present data, the estimated epistatic variance and dominance variance in proportion to additive genetic variance were about 26% and 15%, respectively. The estimates are in the range of those for complex traits reported in previous studies. Additive and non-additive variances are usually estimated using a model with a pedigree-based relationship matrix. In dairy cattle, it was reported that the ratios of dominance variance to additive genetic variance were 118% and 161%, and the ratios of epistatic variance to additive genetic variance were 59% and 2% for days open and service period, respectively, in US Holstein [39]; the ratio of dominance variance to additive genetic variance was 17% for stature in US Holstein [40]; and non-additive variances were, in general, as large as or greater than the additive variances for reproductive traits in Canadian Holstein [2]. In beef cattle, the ratio of dominance variance to additive genetic variance was larger than 50% for weaning weight in Hereford, Gelbvieh and Charolais beef cattle [1,41], and for post-weaning gain in Limousin beef cattle [42]. In pigs, significant contributions of non-additive genetic variance have been reported. The ratios of dominance variance to additive genetic variance ranged from 11% to 31% for reproductive and growth traits in Yorkshire pigs [43]. The ratios were 15% for 21-day litter weight, 44% for number born alive, and 57% for interval between parities in South African Duroc pigs [3]. These results indicate non-additive genetic variations are important for complex traits, especially for fitness traits.

Several previous studies have reported non-additive genetic effects of detected QTL. In chicken, QTL analysis revealed that the non-additive genetic effect was more pronounced prior to 46 days of age, whereas additive genetic effect explained the major portion of the genetic variance later in life [44,45]. In pigs, it was reported that the proportion of non-additive variance relative to the entire QTL variance exceeded 50% in most meat quality and carcass composition traits in a porcine Duroc  $\times$  Pietrain population [46]. Rather than using detected QTLs, the present study estimated additive and non-additive genetic variances using genome-wide SNP markers. This is the first scientific report on estimation of on non-additive genetic variances in livestock using such an approach.

It has been observed that there is a large variation between the estimates of non-additive genetic variances in different studies, which may reflect the different features of various traits and populations. In addition, the large variation could be caused by large sampling error due to insufficient data information. As shown in this study, the standard errors of non-additive genetic variance were large. In fact, the estimated epistatic variance was not statistically significantly different from zero. The relative standard errors for the estimated dominance variance and epistatic variance (here defined as standard error/estimated variance) were 4.5 times and 6.5 times as large as that for additive genetic

variance. Similarly, the relative standard error for dominance variance was 4.1 as large as that for additive genetic variance of stature in US Holstein [40]. These results suggest that a large dataset is needed in order to get accurate estimates of non-additive genetic variances.

### Genomic prediction of daily gain

The current and previous studies have shown that non-additive genetic variance is considerable in complex traits. Therefore, it is expected that a model including non-additive genetic effect would increase the prediction accuracy and improve the unbiasedness. In terms of practical animal breeding, focus is the prediction of breeding value (additive genetic value) or the genetic improvement in next generation. In this study, reliabilities of genomic predicted breeding value using model MAE, MAD and MAED were 0.3%, 0.7% and 1.0% units higher than that using the additive genetic model. In addition the models including non-additive genetic effects slightly improved unbiasedness of genomic prediction.

Compared to large non-additive genetic variances, the gain in reliability of genomic predictions by including non-additive genetic effects in the prediction model was relatively small. The small improvement indicates the difficulty to distinguish additive and non-additive genetic effects. Using traditional BLUP model with pedigree-based relationship matrix, a previous study on stature in US Holsteins [47] found that animals with large amount of additive information were influenced little by the inclusion of dominance, but large for the animals with a large amount of dominance information. The superiority of model MAE over model MA for genomic prediction is negligible, which was consistent with the large standard error for the estimates of epistatic variance. It should be noted that the current study was based on a small dataset. The data information might not be sufficient to distinguish additive genetic and non-additive genetic effects efficiently. It is expected that the benefit for genomic prediction by including non-additive genetic effects in the model will be larger when using larger reference dataset, especially when reference dataset including records of crossbred animals. Moreover, an additive and non-additive genetic model could be benefit for exploiting specific combining ability. A simulation study with various scenarios by [48] reported that the ratio of dominance response to additive genetic selection ranged from 3.8 to 16.6% by one generation of selection.

In a genetic evaluation system, there are two limitations in practice for using a model with both additive and non-additive genetic effects for genomic prediction. The first limitation is that the computational demand for models with both additive and non-additive genetic effects are generally high, since both additive and

non-additive genomic relationship matrices are dense. This requires more powerful computers and/or more efficient algorithms. Second, a reference population often consists of sires which have records of progeny performance, and popularly used response variables are conventional estimated breeding value (EBV), de-regressed EBV or mean of corrected progenies' performances which are more informative than individual observation. However, these pseudo observations are appropriate for an additive genetic model, but not appropriate for a model that includes non-additive genetics effects because the interaction effects observed in the offspring are not related to interactions of genes in the sire. A genomic prediction model including non-additive genetic effects requires that the response variable is individual record. An extension of single-step model [24–26] to handle non-additive genetic effects may be a good approach, because this approach allows predicting breeding value using the observations from both genotyped and non-genotyped animals by combining genomic and pedigree information into a joint relationship matrix.

The current study was based on data from purebred Duroc pigs. Non-additive genetic variation in a crossbred population is expected to be much larger than that in a purebred population. In farm animals such like poultry, pig, and beef cattle, crossbreds are usually the end product. Nonetheless, selection is carried out in the purebreds to improve performance of the crossbreds. The information from crossbred population will be useful for increasing accuracy of genetic evaluation of animals in the purebreds due to additional information of the relatives in crossbred. Moreover, the information of crossbred population allows selecting candidates in purebreds for the performance of their crossbred offspring. According to the results from the present study, when using data of crossbreds, a model including non-additive genetic effects, especially dominance effect, is expected to yield a substantial improvement of genetic evaluations.

It can be concluded that the present method is a feasible approach to estimate additive and non-additive variances and predict genomic breeding values. Non-additive genetic effects are important sources of genetic variation for daily gain in pigs. Genomic prediction models including non-additive genetic effects can improve accuracy and unbiasedness of genomic predicted breeding value.

### Author Contributions

Conceived and designed the experiments: GS OFC MSL. Performed the experiments: GS. Analyzed the data: GS. Contributed reagents/materials/analysis tools: GS TO MH. Wrote the paper: GS OFC MSL.

### References

- Gengler N, VanVleck LD, MacNeil MD, Misztal I, Paricote FA (1997) Influence of dominance relationships on the estimation of dominance variance with sire-dam subclass effects. *Journal of Animal Science* 75: 2885–2891.
- Palucci V, Schaeffer LR, Miglior F, Osborne V (2007) Non-additive genetic effects for fertility traits in Canadian Holstein cattle. *Genetics Selection Evolution* 39: 181–193.
- Norris D, Varona L, Ngambi JW, Visser DP, Mbajjorgu CA, et al. (2010) Estimation of the additive and dominance variances in SA Duroc pigs. *Livestock Science* 131: 144–147.
- Jiang L, Liu J, Sun D, Ma P, Ding X, et al. (2010) Genome wide association studies for milk production traits in Chinese Holstein population. *PLoS One* 5: e13661.
- Hayes B, Goddard M (2010) Genome-wide association and genomic selection in animal breeding. *Genome* 53: 876–883.
- Petterson M, Besnier F, Siegel PB, Carlborg O (2011) Replication and explorations of high-order epistasis using a large advanced intercross line pedigree. *PLoS Genet* 7: e1002180.
- Sahana G, Mailund T, Lund MS, Gulbrandsen B (2011) Local genealogies in a linear mixed model for genome-wide association mapping in complex pedigree populations. *PLoS One* 6: e27061.
- Fu WX, Liu Y, Lu X, Niu XY, Ding XD, et al. (2012) A genome-wide association study identifies two novel promising candidate genes affecting *Escherichia coli* F4ab/F4ac susceptibility in swine. *PLoS One* 7: e32127.
- VanRaden PM, Sullivan PG (2010) International genomic evaluation methods for dairy cattle. *Genetics Selection Evolution* 42: 7.
- Lund MS, de Ross SP, de Vries AG, Druet T, Ducrocq V, et al. (2011) A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol* 43: 43.
- Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: Progress and challenges. *Journal of Dairy Science* 92: 433–443.
- Su G, Gulbrandsen B, Gregersen VR, Lund MS (2010) Preliminary investigation on reliability of genomic estimated breeding values in the Danish Holstein population. *Journal of Dairy Science* 93: 1175–1183.
- Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genetics Research* 91: 47–60.

14. Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
15. VanRaden PM (2008) Efficient Methods to Compute Genomic Predictions. *Journal of Dairy Science* 91: 4414–4423.
16. de los Campos G, Naya H, Gianola D, Crossa J, Legarra A, et al. (2009) Predicting quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182: 375–385.
17. Meuwissen THE (2009) Accuracy of breeding values of 'unrelated' individuals predicted by dense SNP genotyping. *Genetics Selection Evolution* 41: 35.
18. Lund MS, Sahana G, de Koning DJ, Su G, Carlborg Ö (2009) Comparison of analyses of the QTLMAS XII common dataset. I: Genomic selection. *BMC Proceedings* 3: s1.
19. Guo G, Lund MS, Zhang Y, Su G (2010) Comparison between genomic predictions using daughter yield deviation and conventional estimated breeding value as response variables. *Journal of Animal Breeding and Genetics* 127: 423–432.
20. VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD, et al. (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. *Journal of Dairy Science* 92: 16–24.
21. Ostersen T, Christensen OF, Henryon M, Nielsen B, Su GS, et al. (2011) Deregressed EBV as the response variable yield more reliable genomic predictions than traditional EBV in pure-bred pigs. *Genetics Selection Evolution* 43: 38.
22. Goddard M (2009) Genomic selection: prediction of accuracy and maximisation of long term response. *Genetica* 136: 245–257.
23. Strandén I, Garrick DJ (2009) Technical note: Derivation of equivalent computing algorithms for genomic predictions and reliabilities of animal merit. *J Dairy Sci* 92: 2971–2975.
24. Legarra A, Aguilar I, Misztal I (2009) A relationship matrix including full pedigree and genomic information. *Journal of Dairy Science* 92: 4656–4663.
25. Aguilar I, Misztal I, Johnson DL, Legarra A, Tsuruta S, et al. (2010) Hot topic: A unified approach to utilize phenotypic, full pedigree, and genomic information for genetic evaluation of Holstein final score. *Journal of Dairy Science* 93: 743–752.
26. Christensen OF, Lund MS (2010) Genomic prediction when some animals are not genotyped. *Genetics Selection Evolution* 42: 2.
27. Misztal I, Legarra A, Aguilar I (2009) Computing procedures for genetic evaluation including phenotypic, full pedigree, and genomic information. *Journal of Dairy Science* 92: 4648–4655.
28. Henderson CR (1985) Best linear unbiased prediction of nonadditive genetic merits in noninbred populations. *Journal of Animal Science* 60: 7.
29. Gilmour AR, Thompson R, Cullis BR (1995) Average information REML: An efficient algorithm for variance parameter estimation in linear mixed models. *Biometrics* 51: 1440–1450.
30. Dunn OJ, Clark V (1971) Comparison of tests of the equality of dependent correlation coefficients. *Journal of the American Statistical Association* 66: 904–908.
31. Stüger JH (1980) Tests for comparing elements of a correlation matrix. *Psychological Bulletin* 87: 245–251.
32. Su G, Madsen P, Nielsen US, Mäntysaari EA, Aamand GP, et al. (2012) Genomic prediction for Nordic Red Cattle using one-step and selection index blending. *Journal of Dairy Science* 95: 909–917.
33. Harris BL, Johnson DL (2010) Genomic predictions for New Zealand dairy bulls and integration with national genetic evaluation. *Journal of Dairy Science* 93: 1243–1252.
34. Yang JA, Benyamin B, McEvoy BP, Gordon S, Henders AK, et al. (2010) Common SNPs explain a large proportion of the heritability for human height. *Nature Genetics* 42: 565–U131.
35. Visscher PM, Yang JA, Goddard ME (2010) A Commentary on 'Common SNPs Explain a Large Proportion of the Heritability for Human Height' by Yang et al. (2010). *Twin Research and Human Genetics* 13: 517–524.
36. Duthie C, Simm G, Doeschl-Wilson A, Kalm E, Knap PW, et al. (2010) Epistatic analysis of carcass characteristics in pigs reveals genomic interactions between quantitative trait loci attributable to additive and dominance genetic effects. *J Anim Sci* 88: 2219–2234.
37. Xu S, Jia Z (2007) Genomewide analysis of epistatic effects for quantitative traits in barley. *Genetics* 175: 1955–1963.
38. Estelle J, Gil F, Vazquez JM, Latorre R, Ramirez G, et al. (2008) A quantitative trait locus genome scan for porcine muscle fiber traits reveals overdominance and epistasis. *J Anim Sci* 86: 3290–3299.
39. Hoeschele I (1991) Additive and nonadditive genetic variance in female fertility of Holsteins. *Journal of Dairy Science* 74: 1743–1752.
40. Misztal I (1997) Estimation of variance components with large-scale dominance models. *Journal of Dairy Science* 80: 965–974.
41. Duangjinda M, Bertrand JK, Misztal I, Druet T (2001) Estimation of additive and nonadditive genetic variances in Hereford, Gelbvieh, and Charolais by Method R. *Journal of Animal Science* 79: 2997–3001.
42. Gengler N, Misztal I, Bertrand JK, Culbertson MS (1998) Estimation of the dominance variance for postweaning gain in the US Limousin population. *Journal of Animal Science* 76: 2515–2520.
43. Culbertson MS, Mabry JW, Misztal I, Gengler N, Bertrand JK, et al. (1998) Estimation of dominance variance in purebred Yorkshire swine. *Journal of Animal Science* 76: 448–451.
44. Carlborg O, Kerje S, Schutz K, Jacobsson L, Jensen P, et al. (2003) A global search reveals epistatic interaction between QTL for early growth in the chicken. *Genome Research* 13: 413–421.
45. Carlborg R, Hocking PM, Burt DW, Haley CS (2004) Simultaneous mapping of epistatic QTL in chickens reveals clusters of QTL pairs with similar genetic effects on growth. *Genetical Research* 83: 197–209.
46. Grosse-Brinkhaus C, Jonas E, Buschbell H, Phatsara C, Tesfaye D, et al. (2010) Epistatic QTL pairs associated with meat quality and carcass composition traits in a porcine Duroc x Pietrain population. *Genetics Selection Evolution* 42: 39.
47. Varona L, Misztal I, Bertrand JK, Lawlor TJ (1998) Effect of full sibs on additive breeding values under the dominance model for stature in United States Holsteins. *Journal of Dairy Science* 81: 1126–1135.
48. Varona L, Misztal I (1999) Prediction of parental dominance combinations for planned matings, methodology, and simulation results. *Journal of Dairy Science* 82: 2186–2191.