

Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation

Motohide Nishio*, Masahiro Satoh

NARO Institute of Livestock and Grassland Science, Tsukuba, Japan

Abstract

We evaluated the performance of GBLUP including dominance genetic effect (GBLUP-D) by estimating variances and predicting genetic merits in a computer simulation and 2 actual traits (T4 and T5) in pigs. In simulation data, GBLUP-D explained more than 50% of dominance genetic variance. Moreover, GBLUP-D yielded estimated total genetic effects over 1.2% more accurate than those yielded by GBLUP. In particular, when the dominance genetic variance was large, the accuracy could be substantially improved by increasing the number of markers. The dominance genetic variances in T4 and T5 accounted for 9.6% and 6.3% of the phenotypic variances, respectively. Estimates of such small dominance genetic variances contributed little to the improvement of the accuracies of estimated total genetic effects. In both simulation and pig data, there were nearly no differences in the estimates of additive genetic effects or their variance between GBLUP-D and GBLUP. Therefore, we conclude GBLUP-D is a feasible approach to improve genetic performance in crossbred populations with large dominance genetic variation and identify mating systems with good combining ability.

Citation: Nishio M, Satoh M (2014) Including Dominance Effects in the Genomic BLUP Method for Genomic Evaluation. PLoS ONE 9(1): e85792. doi:10.1371/journal.pone.0085792

Editor: Xinping Cui, University of California, Riverside, United States of America

Received: July 21, 2013; **Accepted:** December 6, 2013; **Published:** January 8, 2014

Copyright: © 2014 Nishio, Satoh. 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 authors have no support or funding to report.

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

* E-mail: mtnishio@affrc.go.jp

Introduction

Genomic selection refers to the use of genome-wide dense single nucleotide polymorphism (SNP) markers to predict breeding values and subsequently select individuals [1]. Several approaches of genomic prediction have been presented. One of them is the genomic best linear unbiased prediction (GBLUP), which uses genomic information in the form of a genomic relationship matrix that defines the additive genetic covariance between individuals [2,3]. The genomic relationship coefficients are estimated with higher accuracy than when using pedigree information because genomic information can capture of Mendelian sampling across the genome. GBLUP has become popular approach in genomic selection of dairy cattle [4,5] because it is simple and has low computational requirements [6,7].

Most published models only include additive genetic effects [8], and little research has been performed to expand these models to predict genetic merits to account for dominance genetic effects. It can be argued that such expansion is difficult because calculation becomes complicated and de-regressed estimated breeding values are used as phenotypes in most applications of genomic selection [9]. However, dominance genetic effect is of theoretical and practical important because it is heavily used in crosses of animal breeds. In fact, assortative mating and mate allocation boost the field performances of livestock [10]. Genomic selection has, therefore, renewed the interest in the prediction of dominance genetic effects. For example, the dominance genetic variance accounted for 5.6% of the phenotypic variance by GBLUP including dominance genetic effect [11]. More recently, GBLUP method including dominance genetic effect was suggested and the

software (GVCBLUP) are already available online (<http://animalgene.umn.edu/>) [12–14].

The present study aimed to evaluate the performance of GBLUP including dominance genetic effect by estimating variance components and predicting genetic effects for both simulation and actual pig data.

Materials and Methods

Stochastic Simulation

A historical population was simulated to establish mutation drift equilibrium. The simulated genome comprised one chromosome 1 Morgan long that contained 6,000 SNP markers and 300 randomly spaced biallelic quantitative trait loci (QTL). In the first generation of the historical population, the initial allele frequencies of all markers and QTL were assumed to be 0.5. The recurrent mutation process was applied, and the mutation rate of markers and QTL was 5.0×10^{-4} per locus per generation. Recombinations were sampled from a Poisson distribution with a mean of 1 per Morgan and were then randomly placed along the chromosome. The historical population evolved over 2,000 generations of random mating and random selection with a population size of 100 (50 males and 50 females) to reach mutation–drift balance.

After 2,000 historical generations, a base population (G0) and the subsequent 6 generations (G1 to G6) were generated as a recent population. The population size of G0 increased to 300 (150 males and 150 females). In G1 to G6, 30 sires were randomly selected and mated to 150 dams in each generation. Each dam had 1 son and 1 daughter; thus, each sire had 5 sons and 5 daughters.

In G0, 1,000 markers and 50 QTL were randomly selected among the segregating markers and QTL with minor allele frequencies >0.05. Let Q_{1j} and Q_{2j} be 2 alleles at the j th QTL. The genetic values are then given by α_j , δ_j , and $-\alpha_j$ for genotypes $Q_{1j}Q_{1j}$, $Q_{1j}Q_{2j}$, and $Q_{2j}Q_{2j}$, respectively. The value of α_j was drawn from a gamma distribution with a shape parameter of 0.42; its sign was drawn at random with equal chance. According to the previous simulation study including dominance genetic effects [15], the value of δ_j was determined as the product of the absolute of α_j and the degree of dominance, which was drawn from a normal distribution $N(0, \tau^2)$. The total genetic effect (g_i) of the j th animal was calculated by summing all QTL genotypic values, and its variance (σ_g^2) was calculated as the sum of additive and dominance genetic variances (σ_a^2 and σ_d^2) [16], which were calculated as follows:

$$\sigma_a^2 = \sum_{j=1}^{N_q} 2q_j(1-q_j)\{\alpha_j + (1-2q_j)\delta_j\}^2,$$

$$\sigma_d^2 = \sum_{j=1}^{N_q} \{2q_j(1-q_j)\delta_j\}^2,$$

where q_j is the frequency of Q_{1j} . The broad-sense heritability (h^2) of the trait was 0.3. To obtain phenotypic values, an environmental effect was added to the total genetic effect, which was sampled from a normal distribution $N(0, (1-h^2)\sigma_g^2/h^2)$.

The phenotypes and genotypes of SNP markers were available for 1,500 and 1,800 individuals from G1 to G5 and G1 to G6, respectively. Thus, the reference population with both phenotypes and genotypes comprised 1,500 individuals from G1 to G5, and the test population with only genotypes comprised 300 individuals in G6.

In a standard simulation scenario, τ and the number of markers in G0 (N_m) were set to 0.5 and 1,000, respectively. To investigate the effects of τ and N_m on the performance of the present method, 2 alternative scenarios were simulated in addition to the standard scenario. Three different values of τ (0.25, 0.5, and 1.0) and N_m (200, 1,000, and 5,000) were simulated in the first and second groups, respectively. For all of these alternatives, only the intended parameter differed from the standard scenario. Twenty replicates were simulated for each scenario.

PIC Pig Data

Publicly available data including pedigree, genotypic, and phenotypic information on a single Pig Improvement Company (PIC) nucleus pig line were used (<http://www.g3journal.org/content/2/4/429/Suppl/DC1>). The total number of individuals was 3,512, and all have phenotypes for 2 traits (T4 and T5) and

genotypes available from the PorcineSNP60 chip ($N=64,233$). These phenotypes were already adjusted for environmental fixed effects such as sex, farm, and year of birth [17]. Then, 1,800 individuals were randomly selected from all individuals whose accuracies from the full PIC dataset exceeded 0.8. From these 1,800 individuals, 1,500 were selected from old generations and defined as the reference population while the other 300 individuals were defined as the test population. Genotypes were filtered for minor allele frequencies less than 0.05. The pig data are summarized in Table 1.

Genomic BLUP Model

The GBLUP including additive and dominance genetic effects termed GBLUP-D. The statistical model of GBLUP-D can be expressed as:

$$\mathbf{y} = \mathbf{Xb} + \mathbf{Za} + \mathbf{Zd} + \mathbf{e},$$

where \mathbf{y} is the vector of phenotypes; \mathbf{b} is the vector of fixed effects; \mathbf{a} and \mathbf{d} are the vector of additive and dominance genetic effects of animals; \mathbf{X} and \mathbf{Z} are incident matrices for the fixed effects, additive, and dominance genetic effect, respectively; and \mathbf{e} is the vector of residuals. Additive and dominance genetic effects were assumed to follow normal distributions: $\mathbf{a} \sim N(0, \mathbf{G}\sigma_a^2)$ and $\mathbf{d} \sim N(0, \mathbf{D}\sigma_d^2)$, where \mathbf{G} and \mathbf{D} are additive and dominance genomic relationship matrices, respectively. These matrices describe the relationships between genotyped individuals and can be constructed from the information on genome-wide SNP markers. Let A_{1j} and A_{2j} be 2 alleles at the j th marker locus and p_j be the frequency of A_{2j} . The \mathbf{G} matrix is created as follows [3]:

$$\mathbf{G} = \frac{\mathbf{M}_a \mathbf{M}_a'}{\sum_{j=1}^{N_m} 2p_j(1-p_j)},$$

where \mathbf{M}_a is the $n \times N_m$ matrix (n is a number of individuals) and the element of \mathbf{M}_a for the i th individual at the j th marker is calculated as follows:

$$\mathbf{M}_{a,i,j} = \begin{cases} -2p_j(A_1A_1) \\ 1-2p_j(A_1A_2) \\ 2-2p_j(A_2A_2) \end{cases}$$

Similarly, \mathbf{M}_d is assumed to be the $n \times N_m$ matrix and the element of \mathbf{M}_d for the i th individual at the j th marker can be calculated as:

Table 1. Summary of PIC pig dataset.

Trait	No. SNPs	Number of animals		Phenotype		Accuracy of estimated breeding value from all PIC datasets	
		Reference	Test	Mean	SD	Mean	SD
T4	27,391	1,500	300	-1.125	2.417	0.875	0.048
T5	27,287	1,500	300	44.107	60.315	0.880	0.048

doi:10.1371/journal.pone.0085792.t001

Table 2. Variance component estimates (\pm standard errors) and heritabilities for simulation data with 3 dominance degrees (0.25, 0.5, and 1.0).

Condition	Method	Genetic variance components			Residual variance	Narrow-sense heritability	Broad-sense heritability
		Additive	Dominance				
$\tau = 0.25$	True value	0.259	0.032	0.736	0.267	0.299	
$N_m = 1,000$	GBLUP	0.277 \pm 0.036	–	0.793 \pm 0.022	0.256	–	
	GBLUP-D	0.277 \pm 0.036	0.029 \pm 0.012	0.764 \pm 0.024	0.259	0.286	
$\tau = 0.5$	True value	0.191	0.108	0.736	0.185	0.289	
$N_m = 1,000$	GBLUP	0.205 \pm 0.029	–	0.865 \pm 0.024	0.192	–	
	GBLUP-D	0.208 \pm 0.030	0.066 \pm 0.016	0.797 \pm 0.026	0.194	0.256	
$\tau = 1.0$	True value	0.138	0.147	0.706	0.139	0.290	
$N_m = 1,000$	GBLUP	0.152 \pm 0.024	–	0.857 \pm 0.024	0.151	–	
	GBLUP-D	0.152 \pm 0.024	0.080 \pm 0.016	0.773 \pm 0.025	0.151	0.231	

doi:10.1371/journal.pone.0085792.t002

$$\mathbf{M}_{d_i,j} = \begin{cases} -2p_j^2(A_1A_1) \\ 2p_j(1-p_j)(A_1A_2) \\ -2(1-p_j)^2(A_2A_2) \end{cases}$$

$$\mathbf{V}(\mathbf{d}) = \mathbf{M}_d \mathbf{M}_d' \sigma_\delta^2 = \mathbf{D} \sigma_d^2,$$

$$\sigma_d^2 = \sum_{j=1}^{N_m} \{2p_j(1-p_j)\}^2 \sigma_\delta^2.$$

This element describes the coefficients of δ_j in dominance deviations. Therefore, \mathbf{d} and its variance can be derived as follows:

$$\mathbf{d} = \mathbf{M}_d \boldsymbol{\delta},$$

where $\boldsymbol{\delta}$ is the N_m dimensional vector of the j th element, which is δ_j . This dominance formula was also used in GVCBLUP [14]. Assuming the dominance genetic effects at different marker loci are identically and independently distributed normal variables, the variance of the genome-wide dominance effect is calculated as follows:

Consequently, \mathbf{D} can be calculated using \mathbf{M}_d :

$$\mathbf{D} = \frac{\mathbf{M}_d \mathbf{M}_d'}{\sum_{j=1}^{N_m} \{2p_j(1-p_j)\}^2}.$$

Variance components were estimated with average information restricted maximum likelihood (REML) [18]. The dataset of reference population were used to predict genetic effects of the

Table 3. Variance component estimates (\pm standard errors) and heritabilities for simulation data with 200, 1,000, and 5,000 markers.

Condition	Method	Genetic variance components			Residual variance	Narrow-sense heritability	Broad-sense heritability
		Additive	Dominance				
$\tau = 0.5$	True value	0.195	0.103	0.739	0.188	0.287	
$N_m = 200$	GBLUP	0.158 \pm 0.026	–	0.878 \pm 0.024	0.153	–	
	GBLUP-D	0.158 \pm 0.026	0.048 \pm 0.012	0.831 \pm 0.024	0.152	0.199	
$\tau = 0.5$	True value	0.191	0.108	0.736	0.185	0.289	
$N_m = 1,000$	GBLUP	0.205 \pm 0.029	–	0.865 \pm 0.024	0.192	–	
	GBLUP-D	0.208 \pm 0.030	0.066 \pm 0.016	0.797 \pm 0.026	0.194	0.256	
$\tau = 0.5$	True value	0.211	0.075	0.685	0.217	0.295	
$N_m = 5,000$	GBLUP	0.206 \pm 0.027	–	0.761 \pm 0.021	0.213	–	
	GBLUP-D	0.207 \pm 0.028	0.063 \pm 0.016	0.698 \pm 0.024	0.214	0.279	

doi:10.1371/journal.pone.0085792.t003

Table 4. Accuracies of estimates ($r_{\hat{a},a}$, $r_{\hat{d},d}$, and $r_{\hat{g},g}$) and regression coefficients of estimates on their true values ($b_{\hat{a},a}$, $b_{\hat{d},d}$, and $b_{\hat{g},g}$) in the test population for simulation data with 3 dominance degrees (0.25, 0.5, and 1.0).

Condition	Method	$r_{\hat{a},a}$	$r_{\hat{d},d}$	$r_{\hat{g},g}$	$b_{\hat{a},a}$	$b_{\hat{d},d}$	$b_{\hat{g},g}$
$\tau = 0.25$	GBLUP	0.803	–	0.760	0.898	–	0.939
$N_m = 1,000$	GBLUP-D	0.804	0.212	0.769	0.902	0.609	0.942
$\tau = 0.5$	GBLUP	0.743	–	0.616	0.891	–	0.976
$N_m = 1,000$	GBLUP-D	0.745	0.339	0.664	0.900	0.893	0.994
$\tau = 1.0$	GBLUP	0.711	–	0.466	1.001	–	0.937
$N_m = 1,000$	GBLUP-D	0.712	0.478	0.581	1.006	1.189	1.035

doi:10.1371/journal.pone.0085792.t004

genotyped individuals in test population. The GBLUP-D model solutions yielded the estimates of additive ($\hat{\mathbf{a}}$) and dominance genetic effects ($\hat{\mathbf{d}}$). The estimates of the total genetic effect ($\hat{\mathbf{g}}$) were calculated by the sum of $\hat{\mathbf{a}}$ and $\hat{\mathbf{d}}$. In GBLUP, $\hat{\mathbf{g}}$ equals $\hat{\mathbf{a}}$, because the dominance genetic effect is not considered. The predictive ability of the model was evaluated from accuracy and unbiasedness of estimates in the test population. The accuracies of estimated additive, dominance, and total genetic effects ($r_{\hat{a},a}$, $r_{\hat{d},d}$, and $r_{\hat{g},g}$, respectively) were measured as the correlations between the estimates and true values. Unbiasedness ($b_{\hat{a},a}$, $b_{\hat{d},d}$, and $b_{\hat{g},g}$) was measured using the regressions of estimates on true values. A regression coefficient of one denotes unbiasedness. Since the true values are unknown in pig data, estimated breeding values from the full PIC dataset (a^*) and phenotypes (y) were used instead of true additive and total genetic effects, respectively. In real data, the predictive ability of dominance genetic effect cannot be calculated and was inferred from that of total genetic effect.

Results

Stochastic Simulation

Tables 2 and 3 show the estimates of variance components and heritability in simulation data with various values of τ and N_m . For all values of τ and N_m , there were nearly no differences in estimates of additive genetic variance and narrow-sense heritability between GBLUP and GBLUP-D. The ratios of dominant genetic variance estimated by GBLUP-D were 90.6%, 61.1%, and 54.4% with a τ of 0.25, 0.5, and 1.0 and 46.6%, 61.1%, and 84.0% with an N_m of 200, 1,000, and 5,000, respectively.

Tables 4 and 5 show the accuracies and unbiasedness of estimated genetic values calculated by GBLUP and GBLUP-D in simulation data. For all values of τ and N_m , $r_{\hat{a},a}$ and $b_{\hat{a},a}$ were almost equal between GBLUP-D and GBLUP. In GBLUP-D, $r_{\hat{d},d}$ and $b_{\hat{d},d}$ increased with increasing τ and N_m . Meanwhile, $r_{\hat{g},g}$ values in GBLUP-D exceeded those in GBLUP by 1.2%, 7.8%, and 24.7% with a τ of 0.25, 0.5, and 1.0 and by 1.9%, 7.8%, and 4.2% with an N_m of 200, 1,000, and 5,000, respectively. In GBLUP-D, larger values of τ resulted in $b_{\hat{g},g}$ being closer to 1.

PIC Pig Data

In T4 and T5 from the pig data, dominance genetic variance accounted for 9.6% and 6.3% of the phenotypic variance, respectively (Table 6). The estimated additive genetic variances and residual variances calculated by GBLUP-D were smaller than those calculated by GBLUP. Thus, GBLUP-D consequently

Table 5. Accuracies of estimates ($r_{\hat{a},a}$, $r_{\hat{d},d}$, and $r_{\hat{g},g}$) and regression coefficients of estimates on their true values ($b_{\hat{a},a}$, $b_{\hat{d},d}$, and $b_{\hat{g},g}$) in the test population for simulation data with 200, 1,000, and 5,000 markers.

Condition	Method	$r_{\hat{a},a}$	$r_{\hat{d},d}$	$r_{\hat{g},g}$	$b_{\hat{a},a}$	$b_{\hat{d},d}$	$b_{\hat{g},g}$
$\tau = 0.5$	GBLUP	0.689	–	0.552	1.049	–	1.027
$N_m = 200$	GBLUP-D	0.696	0.246	0.563	1.062	0.547	0.938
$\tau = 0.5$	GBLUP	0.743	–	0.616	0.891	–	0.976
$N_m = 1,000$	GBLUP-D	0.745	0.339	0.664	0.900	0.893	0.994
$\tau = 0.5$	GBLUP	0.799	–	0.694	1.004	–	1.035
$N_m = 5,000$	GBLUP-D	0.801	0.374	0.723	1.010	0.961	1.032

doi:10.1371/journal.pone.0085792.t005

yielded lower narrow-sense heritability and higher broad-sense heritability than GBLUP.

In T4 and T5 from the pig data, there were nearly no differences between GBLUP and GBLUP-D with respect to $r_{\hat{a},a}$ (Table 7). In T5, $r_{\hat{g},y}$ and $b_{\hat{g},y}$ were slightly higher in GBLUP-D than GBLUP, whereas $r_{\hat{d},y}$ and $b_{\hat{d},y}$ were not higher in GBLUP-D in T4.

Discussion

Stochastic Simulation

The GBLUP-D method captured the substantial ratios of the dominance genetic variances and estimated the individual dominance genetic effects although there were nearly no differences in estimates of additive genetic variance and narrow-sense heritability between GBLUP and GBLUP-D. Our result indicates that GBLUP-D is expected to improve performance of the crossbreds, in particular when degree of dominance is large.

In the present study, the simulated genome comprised one chromosome 1 Morgan long. However, the whole genome sizes of livestock are larger. Here, another simulation data were constructed to evaluate the effect of the genome size on predictive ability. This simulation data comprised five chromosomes of 1 Morgan. The numbers of markers and QTL set to be 5,000 and 250 to obtain the same distances between markers and QTL as the initial simulation. Table 8 shows the accuracies and unbiasedness of estimated genetic values in this simulation data. The dominance genetic effects could be captured in this data, but the accuracies of additive, dominance and total genetic effects decreased in comparison with the genome of 1 chromosome. Hence, large size of reference population would be required when genome size is large.

PIC Pig Data

The degrees of estimated dominance genetic variance in T4 and T5 were nearly equal to those in simulation data with a τ of 0.25. When τ was 0.25, $r_{\hat{g},g}$ and $b_{\hat{g},g}$ were 1.2% and 0.3% higher in GBLUP-D than GBLUP. These results indicate the predictive ability of GBLUP-D in T4 and T5 only improved slightly because the degrees of dominance genetic variance were too small. In fact, $r_{\hat{g},y}$ and $b_{\hat{g},y}$ in GBLUP-D were little improved in comparison with GBLUP. In general, the degree of dominance genetic variance is expected to be much larger in crossbred populations than purebred ones. Since the present study was based on data from purebred PIC pig data, the degrees of dominance genetic variance in T4 and T5 might have been small.

Table 6. Variance components estimates (\pm standard errors) and heritabilities for PIC pig data.

Genetic variance components						
Trait	Method	Additive	Dominance	Residual variance	Narrow-sense heritability	Broad-sense heritability
T4	GBLUP	1.909 \pm 0.189	–	3.678 \pm 0.138	0.342	–
	GBLUP-D	1.735 \pm 0.201	0.537 \pm 0.219	3.331 \pm 0.185	0.310	0.405
T5	GBLUP	1298.1 \pm 123.2	–	2198.2 \pm 84.3	0.371	–
	GBLUP-D	1239.0 \pm 129.3	220.8 \pm 125.9	2049.3 \pm 113.1	0.353	0.416

doi:10.1371/journal.pone.0085792.t006

Practical Use of GBLUP-D

GBLUP-D has two practical uses. First, selection on the basis of GBLUP-D in crossbreds is useful for commercial production. In swine and poultry, crossbreds are the end product. The marker information from a purebred and its crossbred relatives enables the selection of candidate purebreds for the performance of their crossbred offspring [11]. Second, GBLUP-D could allow mating allocation to exploit dominance. An extra response will be obtained when an appropriate design of future matings using mating allocation techniques is implemented [10,19].

Additive and Dominance Relationship Matrix

In the presence of dominance genetic effects, the breeding values of $A_{1j}A_{1j}$, $A_{1j}A_{2j}$, and $A_{2j}A_{2j}$ at the j th locus are $-2p_j\{\alpha_j + (1 - 2p_j)\delta_j\}$, $(1 - 2p_j)\{\alpha_j + (1 - 2p_j)\delta_j\}$, and $(2 - 2p_j)\{\alpha_j + (1 - 2p_j)\delta_j\}$, respectively. The additive genomic relationship matrix in GBLUP-D should be constructed considering δ_j . If all QTL are of complete dominance, then δ_j is 1. However, in practice, the value of δ_j cannot be determined because the degree of dominance is unknown. Therefore, in the present study, the additive genomic relationship matrix was the same in GBLUP-D and GBLUP. Although there are nearly no differences in the estimates of additive genetic effects or their variance between GBLUP and GBLUP-D, the additive genomic relationship matrix including δ may yield good estimates of them.

A GBLUP method including dominance genetic effects was also proposed in a previous study [11]. In that model, the additive genomic relationship matrix is same as that in traditional GBLUP. However, the dominance genomic relationship matrix (\mathbf{D}^*) in the previous study differs from that in the present study and is defined as follows:

$$\mathbf{D}^* = \frac{\mathbf{H}\mathbf{H}'}{\sum_{j=1}^{N_m} 2p_j(1-p_j)\{1-2p_j(1-p_j)\}}$$

where the element of \mathbf{H} for the i th individual at the j th marker is calculated as

$$\mathbf{H}_{i,j} = \begin{cases} -2p_j(1-p_j)(A_1A_1) \\ 1-2p_j(1-p_j)(A_1A_2) \\ -2p_j(1-p_j)(A_2A_2) \end{cases}$$

This element corresponds to the heterozygosity coefficient but not the dominance genetic effect. To compare the performance of GBLUP-D in the present study and the model in the previous study [11], predictive ability was calculated in the standard simulation scenario (Table 9). GBLUP-D yielded higher accuracies of additive, dominance, and total genetic effects than the previous model [11]. This might be because the heterozygosity coefficient includes part of the additive genetic effect. In fact, the previous study [12] reported that dominance genetic variance calculated from the previous model [11] was larger than that from GBLUP-D.

Assuming linkage equilibrium and uncorrelated marker effects, the dominance genetic variances in the present study (σ_d^2) and the previous study [11] ($\sigma_{d^*}^2$) are calculated as follows:

Table 7. Accs of estimates ($r_{\hat{a},a}$ and $r_{\hat{g},y}$) and regression coefficients ($b_{\hat{a},a}$ and $b_{\hat{g},y}$) of \hat{a} on full PIC dataset (a^*) and \hat{g} on phenotypic value (y) in the test population for the PIC pig dataset.

Trait	Method	$r_{\hat{a},a^*}$	$r_{\hat{g},y}$	$b_{\hat{a},a^*}$	$b_{\hat{g},y}$
T4	GBLUP	0.455	0.286	0.637	0.726
	GBLUP-D	0.456	0.286	0.710	0.724
T5	GBLUP	0.379	0.288	0.631	0.786
	GBLUP-D	0.376	0.299	0.654	0.805

doi:10.1371/journal.pone.0085792.t007

Table 8. Accuracies of estimates ($r_{\hat{a},a}$, $r_{\hat{d},d}$, and $r_{\hat{g},g}$) and regression coefficients of estimates on their true values ($b_{\hat{a},a}$, $b_{\hat{d},d}$, and $b_{\hat{g},g}$) in the test population for simulation data with 3 dominance degrees (0.25, 0.5, and 1.0) when genome comprises of 5 chromosomes with 1 Morgan each.

Condition	Method	$r_{\hat{a},a}$	$r_{\hat{d},d}$	$r_{\hat{g},g}$	$b_{\hat{a},a}$	$b_{\hat{d},d}$	$b_{\hat{g},g}$
$\tau = 0.25$	GBLUP	0.672	–	0.636	1.012	–	1.017
$N_m = 1,000$	GBLUP-D	0.673	0.148	0.641	1.013	0.810	1.018
$\tau = 0.5$	GBLUP	0.646	–	0.502	1.015	–	1.018
$N_m = 1,000$	GBLUP-D	0.647	0.244	0.528	1.016	1.011	1.015
$\tau = 1.0$	GBLUP	0.612	–	0.490	1.112	–	1.124
$N_m = 1,000$	GBLUP-D	0.612	0.348	0.525	1.122	1.030	1.120

doi:10.1371/journal.pone.0085792.t008

Table 9. Accuracies of estimates ($r_{a,a}$, $r_{d,d}$, and $r_{g,g}$) and regression coefficients of estimates on their true values ($b_{a,a}$, $b_{d,d}$, and $b_{g,g}$) in the test population for simulation data ($\tau = 0.5, N_m = 1,000$).

Method	$r_{a,a}$	$r_{d,d}$	$r_{g,g}$	$b_{a,a}$	$b_{d,d}$	$b_{g,g}$
GBLUP	0.743	–	0.616	0.891	–	0.976
GBLUP-D	0.745	0.339	0.664	0.900	0.893	0.994
Su et al. [17]	0.709	0.239	0.655	0.939	0.569	0.967

doi:10.1371/journal.pone.0085792.t009

$$\sigma_d^2 = \sum_{j=1}^{N_m} \{2p_j(1-p_j)\}^2 \sigma_\delta^2,$$

$$\sigma_{d^*}^2 = \sum_{j=1}^{N_m} 2p_j(1-p_j)\{1-2p_j(1-p_j)\} \sigma_\delta^2.$$

In addition, the estimated dominance genetic effects in the present study ($\hat{\mathbf{d}}$) and the previous study [11] ($\hat{\mathbf{d}}^*$) are calculated as follows:

$$\hat{\mathbf{d}} = \sigma_d^2 \mathbf{DZ}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}),$$

$$\hat{\mathbf{d}}^* = \sigma_{d^*}^2 \mathbf{D}^*\mathbf{Z}'\mathbf{V}^{-1}(\mathbf{y} - \mathbf{Xb}),$$

References

- Muewissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic effect using genome-wide dense marker maps. *Genetics* 157: 1819–1829.
- Hayes BJ, Bowman RJ, Chamberlain AL, Goddard ME (2009) Invited review: Genomic selection in dairy cattle: progress and challenges. *J Dairy Sci* 92: 433–443.
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91: 4414–4423.
- McHugh N, Muewissen TH, Cromie AR, Sonesson AK (2011) Use of female information in dairy cattle genomic breeding programs. *J Dairy Sci* 94: 4109–4118.
- Wiggans GR, VanRaden PM, Cooper T. 2011. The genomic evaluation system in the United States: Past, present, future. *Journal of Dairy Science* 94, 3202–3211.
- VanRaden PM, Van Tassell CP, Wiggans GR, Sonstegard TS, Schnabel RD (2009) Invited review: Reliability of genomic predictions for North American Holstein bulls. *J Dairy Sci* 92: 16–24.
- Hayes BJ, Visscher PM, Goddard ME (2009) Increased accuracy of artificial selection by using the realized relationship matrix. *Genet Res* 91: 47–60.
- Calus MPL (2010) Genomic breeding value prediction: methods and procedures. *Animal* 4: 157–164.
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41: 55.
- Toro MA, Varona L (2010) A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol* 42: 33.
- Su G, Christensen OF, Henryon M, Lund MS (2012) Estimating additive and non-additive genetic variances and prediction genetic merits using genome-wide dense single nucleotide polymorphism markers. *PLoS One* 7: e45293.
- Vitezica ZG, Varona L, Legarra A (2013) On the additive and dominance variance and covariance of individuals within the genomic selection scope. *Genetics*: In press.
- Da Y, Wang S (2013) Mixed model methods for genomic prediction and estimation of variance components of additive and dominance effects using SNP markers. *J Dairy Sci* 96 (E-Suppl.1): 620.
- Wang C, Prakapenka D, Wang S, Runesha HB, Da Y (2013) GVCBLUP 2.1: A computing package for genomic prediction and estimation of variance components for additive and dominance effects using SNP markers. *J Dairy Sci* 96 (E-Suppl.1): 620.
- Wittenburg D, Melzer N, Reinsch N (2011) Including non-additive genetic effects in Bayesian methods for the prediction of genetic values based on genome-wide markers. *BMC Genet* 12: 74.
- Falconer DS, Mackay TFC (1996) Introduction to quantitative genetics. Essex: Longman Group. 126 p.
- Cleveland MA, Hickey JM, Forni S (2012) A common dataset for genomic analysis of livestock populations. *G3* 2: 429–435.
- Johnson DL, Thompson R (1995) Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J Dairy Sci* 78: 449–456.
- Sun C, VanRaden PM, O'Connell JR, Weigel KA, Gianola D (2013) Mating programs including genomic relationships and dominance effects. *J Dairy Sci*: In press.
- Schadt EE, Lamb J, Yang X, Zhu J, Edwards S (2005) An interactive genomics approach to infer causal associations between gene expression and disease. *Nat Genet* 37: 710–717.
- Evans DM, Marchini J, Morris AP, Cardon LR (2006) Two-stage two-locus models in genome-wide association. *PLoS One* 2: e157.
- Marchini J, Donnelly P, Cardon LR (2005) Genome-wide strategies for detecting multiple loci that influence complex diseases. *Nat Genet* 37: 413–417.
- Lee SH, van der Werf JHJ, Hayes BJ, Goddard ME, Visscher PM (2008) Predicting unobserved phenotypes for complex traits from whole-genome SNP data. *PLoS Genet* 4: e1000231.
- Xu S (2013) Mapping quantitative trait loci by controlling polygenic background effects. *Genetics*: In press.
- Gianola D, Fernando RL, Stella A (2006) Genomic-assisted prediction of genetic value with semiparametric procedure. *Genetics* 173: 1761–1776.
- Gonzalez-Recio O, Gianola D, Long N, Weigel KA, Rosa GJM (2008) Nonparametric methods for incorporating genomic information into genetic evaluations: an application to mortality in broilers. *Genetics* 178: 2305–2313.
- Yao C, Spurlock DM, Armentano LE, Page Jr CD, VandeHarr JM, et al. (2013) Random forests approach for identifying additive and epistatic single nucleotide polymorphisms associated with residual feed intake in dairy cattle. *J Dairy Sci* 96: 6716–6729.

where \mathbf{V} is the variance of \mathbf{y} . If the distribution of the allelic frequencies is available, $\hat{\mathbf{d}}^*$ can be transformed to $\hat{\mathbf{d}}$.

Epistasis

Increasing knowledge about biological pathways and gene networks highlights the importance of gene–gene interactions, i.e., epistasis; some authors argue that much of the genetic variance in a population is due to such interactions [20–22]. When considering second-order epistasis in GBLUP, the epistatic genomic relationship matrix can be approximately calculated from the Hadamard product of the genomic relationship matrix. For example, additive by additive and additive by dominance interactions are represented as $\mathbf{G}\#\mathbf{G}$ and $\mathbf{G}\#\mathbf{D}$, respectively. The present study tried to use linear mixed models including $\mathbf{G}\#\mathbf{G}$ and $\mathbf{G}\#\mathbf{D}$ for T4 and T5 in pig data. However, the variance components of these epistasis could not be estimated because they were outside parameter space (data not shown). The previous study [11] also included epistatic genomic relationship matrices in GBLUP. In that model, the estimated epistatic variances were almost 0 for daily gain in Danish Duroc pigs. The estimated epistatic variances in the Bayesian model for the percentage of CD8⁺ cells in publicly available mouse data were almost 0 [23]. Although these studies could not detect epistatic effect, in recent, the marker-generated kinship matrices were suggested in a new mixed model method [24] and nonparametric approaches and machine-learning techniques were recommended to model more complex gene interaction patterns [25–27].

Author Contributions

Conceived and designed the experiments: MN MS. Performed the experiments: MN. Analyzed the data: MN. Contributed reagents/materials/analysis tools: MN. Wrote the paper: MN.