



# **Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation**

Motohide Nishio, Masahiro Satoh

## **► To cite this version:**

Motohide Nishio, Masahiro Satoh. Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation. *Genetics Selection Evolution*, 2015, 47 (1), pp.32. <10.1186/s12711-015-0091-y>. <hal-01303965>

**HAL Id: hal-01303965**

**<https://hal.science/hal-01303965v1>**

Submitted on 18 Apr 2016

**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.



HAL Authorization

RESEARCH

Open Access

# Genomic best linear unbiased prediction method including imprinting effects for genomic evaluation

Motohide Nishio\* and Masahiro Satoh

## Abstract

**Background:** Genomic best linear unbiased prediction (GBLUP) is a statistical method used to predict breeding values using single nucleotide polymorphisms for selection in animal and plant breeding. Genetic effects are often modeled as additively acting marker allele effects. However, the actual mode of biological action can differ from this assumption. Many livestock traits exhibit genomic imprinting, which may substantially contribute to the total genetic variation of quantitative traits. Here, we present two statistical models of GBLUP including imprinting effects (GBLUP-I) on the basis of genotypic values (GBLUP-I1) and gametic values (GBLUP-I2). The performance of these models for the estimation of variance components and prediction of genetic values across a range of genetic variations was evaluated in simulations.

**Results:** Estimates of total genetic variances and residual variances with GBLUP-I1 and GBLUP-I2 were close to the true values and the regression coefficients of total genetic values on their estimates were close to 1. Accuracies of estimated total genetic values in both GBLUP-I methods increased with increasing degree of imprinting and broad-sense heritability. When the imprinting variances were equal to 1.4% to 6.0% of the phenotypic variances, the accuracies of estimated total genetic values with GBLUP-I1 exceeded those with GBLUP by 1.4% to 7.8%. In comparison with GBLUP-I1, the superiority of GBLUP-I2 over GBLUP depended strongly on degree of imprinting and difference in genetic values between paternal and maternal alleles. When paternal and maternal alleles were predicted (phasing accuracy was equal to 0.979), accuracies of the estimated total genetic values in GBLUP-I1 and GBLUP-I2 were 1.7% and 1.2% lower than when paternal and maternal alleles were known.

**Conclusions:** This simulation study shows that GBLUP-I1 and GBLUP-I2 can accurately estimate total genetic variance and perform well for the prediction of total genetic values. GBLUP-I1 is preferred for genomic evaluation, while GBLUP-I2 is preferred when the imprinting effects are large, and the genetic effects differ substantially between sexes.

## Background

Genomic imprinting is an epigenetic process that involves DNA methylation and histone modifications that distinguish the expression of maternal and paternal alleles [1]. The expression of an imprinted gene depends on the parent from which it is inherited. Complete inactivation of an imprinted gene results in functional haploidy, with only one of the two copies of the gene expressed. Well known examples of such imprinted genes are *IGF2* (*insulin-like growth factor 2*) in pigs [2] and the *Callipyge* gene in sheep

[3]. Moreover, imprinting may not entail the complete inactivation of a gene. In a study of peripheral blood leukocytes in humans, four of 38 cases exhibited substantial biallelic expression of *IGF2*, although the product level of the maternally-derived gene was lower than that of the paternally-derived gene in all cases [4]. Over 70 imprinted genes have been identified in mice [5], and 24 genes with parent-of-origin effects in beef cattle [6]. Furthermore, quantitative traits such as carcass composition, growth, teat number, and litter size have been suggested to exhibit imprinting effects [7-10]. Thus, imprinting effects may substantially contribute to the total genetic variation of quantitative traits.

\* Correspondence: [mtnishio@affrc.go.jp](mailto:mtnishio@affrc.go.jp)  
NARO Institute of Livestock and Grassland Science, 305-0901, Ikenodai 2  
Tsukuba, Japan

There are several statistical methods for modeling imprinting effects. Using a mixed model, Schaeffer et al. [11] replaced the numerator relationship matrix with a gametic relationship matrix to calculate the expectation of covariance among relatives with imprinting. Essl and Voith [12] suggested that sire and dam models should be constructed separately to assess differences between paternal and maternal imprinting. Neugebauer et al. [13,14] recently fitted a model with correlated paternal and maternal gametes to simultaneously estimate imprinting variances between sexes in pigs and beef cattle. These methods are based on the traditional best linear unbiased prediction (BLUP) method, which uses only pedigree information. More recently, the genomic BLUP (GBLUP) method was developed by modifying the BLUP method to incorporate single nucleotide polymorphism (SNP) information in the form of a genomic relationship matrix that defines the additive genetic covariance among individuals. GBLUP includes genomic information into breeding value estimation and has been used for genomic selection in dairy cattle [15–18]. Therefore, modeling genetic effects by including imprinting effects is expected to improve the predictive ability of GBLUP. Thus, the objectives of this study were twofold: (1) develop a GBLUP method including imprinting effects (termed GBLUP-I hereafter) and (2) estimate genetic variances and assess the accuracies and unbiasedness of genomic predictions using simulation data with varying degrees of imprinting.

## Methods

### Genetic model

Spencer [19] first extended the standard two-allele one locus model of quantitative genetics to account for imprinting. Following the approach of Spencer [19], consider an autosomal biallelic locus with alleles  $A_1$  and  $A_2$  at frequencies  $1-q$  and  $q$ , respectively, in the population. Allele frequencies of males and females were assumed to be the same and under Hardy-Weinberg equilibrium. By denoting a genotype,  $A_iA_j$ ,  $A_i$  and  $A_j$  are the paternally- and maternally-derived alleles, respectively. Following the approach of Spencer [19], the genotypic values for genotypes  $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_1$ , and  $A_2A_2$  are then given by  $a$ ,  $d_1$ ,  $d_2$ , and  $-a$ , respectively. In this study, the mean of two heterozygotes and the difference between two heterozygotes were defined as  $\delta$  and  $\varepsilon$ :

$$\delta = \frac{d_1 + d_2}{2}$$

and

$$\varepsilon = \frac{d_1 - d_2}{2}$$

In this model, the heterozygous genotypic values were  $+\varepsilon$  and  $-\varepsilon$  deviations from  $\delta$ , i.e.,  $d_1$  and  $d_2$  can be

rewritten as  $\delta + \varepsilon$  and  $\delta - \varepsilon$ , respectively (Figure 1). With imprinting, reciprocal heterozygotes differ in their genotypes. For example, in the case of complete inactivation of the maternal allele (i.e.,  $\varepsilon = a$  and  $\delta = 0$ ), the genotypic value of  $A_2A_1$  is the same as that of  $A_2A_2$ , whereas the genotypic value of  $A_1A_2$  is the same as that of  $A_1A_1$ . If the paternally-derived  $A_1$  allele randomly combines with maternally-derived alleles from a population, the frequencies of the genotypes produced will be  $1-q$  for  $A_1A_1$  and  $q$  for  $A_1A_2$ . The genotypic values of  $A_1A_1$  and  $A_1A_2$  are  $a$  and  $d_1$ , respectively. Taking in account the proportions at which they occur, the mean value of genotypes produced from the paternally-derived  $A_1$  allele is  $(1-q)a + qd_1$ . The mean genotypic value in the entire population ( $\mu$ ) is as follows:

$$\begin{aligned}\mu &= (1-q)^2 \cdot a + (1-q)q \cdot d_1 + (1-q)q \cdot d_2 + q^2 \cdot (-a) \\ &= (1-2q)a + 2(1-q)q\delta.\end{aligned}$$

Thus, the average effect of the paternally-derived  $A_1$  allele is calculated from the difference between the mean value of the genotypes produced and population mean as follows:

$$\begin{aligned}(1-q)a + qd_1 - \{(1-2q)a + 2(1-q)q\delta\} \\ = q\{a + (2q-1)\delta + \varepsilon\} = q\alpha_m,\end{aligned}$$

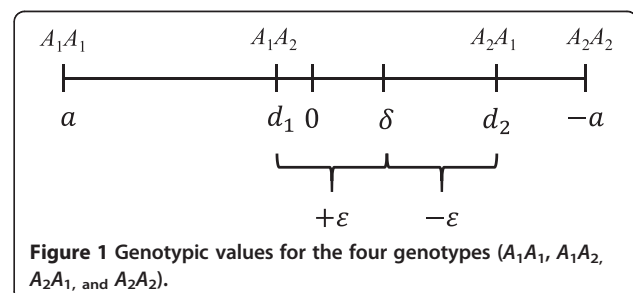
where  $\alpha_m$  is the average effect of the allele substitution in the paternal gamete and is equivalent to the male breeding value of Spencer [19]. Similarly, the average effect of the maternally-derived  $A_1$  allele is as follows:

$$q\{a + (2q-1)\delta - \varepsilon\} = q\alpha_f,$$

where  $\alpha_f$  is the average effect of the allele substitution in the maternal gamete and is equivalent to the female breeding values of Spencer [19]. The average effects of all alleles are in Table 1.

The genotypic deviation of a particular genotype can be calculated from the difference between its genotypic value and the population mean. For example, the genotypic deviation of  $A_1A_2$  is as follows:

$$\begin{aligned}d_1 - \mu &= (2q-1)a + \{1-2(1-q)q\}\delta + \varepsilon \\ &= (2q-1)a + 2(1-q)q\delta + \varepsilon,\end{aligned}$$



where  $\alpha = a + (2q - 1)\delta$ . When there is no imprinting ( $d_1 = d_2 = \delta$  and  $\varepsilon = 0$ )  $\alpha$  is the same as the average effect of the allele substitution [20] and  $(2q - 1)\alpha$  and  $2(1 - q)q\delta$  are the same as the breeding value and dominance deviation of the traditional genetic model. By using  $\delta$  and  $\varepsilon$ , a genotypic deviation can be divided into three terms. Under imprinting, the breeding values and dominance deviations are no longer uncorrelated, which means that the total genetic variance cannot be partitioned into the usual additive and dominance variance [19]. Therefore, in this study, total genetic variance was partitioned into three variances corresponding to  $\alpha$ ,  $\delta$ , and  $\varepsilon$  ( $\sigma_{a'}^2$ ,  $\sigma_{d'}^2$ , and  $\sigma_{i'}^2$ ) as follows:

$$\begin{aligned}\sigma_g^2 &= 2(1-q)q\alpha^2 + \{2q(1-q)\}^2\delta^2 + 2(1-q)q\varepsilon^2 \\ &= \sigma_{a'}^2 + \sigma_{d'}^2 + \sigma_{i'}^2.\end{aligned}$$

When there is no imprinting,  $\sigma_{a'}$  and  $\sigma_{d'}$  are the same as the additive and dominance genetic variance, respectively. In this case, the covariance between the  $\alpha$  and  $\delta$  terms ( $\sigma_{a'd'}$ ) is equal to 0, as follows:

$$\begin{aligned}\sigma_{a'd'} &= (1-q)^2 \cdot 2q\alpha \cdot (-2q^2\delta) \\ &+ (1-q)q \cdot (2q-1)\alpha \cdot 2(1-q)q\delta \\ &+ (1-q)q \cdot (2q-1)\alpha \cdot 2(1-q)q\delta \\ &+ q^2 \cdot -2(1-q)\alpha \cdot \{-2(1-q)^2\delta\} = 0.\end{aligned}$$

The covariance between the  $\alpha$  and  $\varepsilon$  terms ( $\sigma_{a'i'}$ ) is also equal to 0, as follows:

$$\begin{aligned}\sigma_{a'i'} &= (1-q)^2 \cdot 2q\alpha \cdot 0 \\ &+ (1-q)q \cdot (2q-1)\alpha \cdot \varepsilon \\ &+ (1-q)q \cdot (2q-1)\alpha \cdot (-\varepsilon) \\ &+ q^2 \cdot -2(1-q)\alpha \cdot 0 = 0.\end{aligned}$$

Similarly, the covariance between the  $\delta$  and  $\varepsilon$  terms is also equal to 0.

Alternatively, paternal and maternal gametic variances ( $\sigma_{pat}^2$  and  $\sigma_{mat}^2$ , respectively) can be calculated from the

variances of the average effects of paternally- and maternally-derived alleles:

$$\begin{aligned}\sigma_{pat}^2 &= (1-q) \cdot (q\alpha_m)^2 + q \cdot \{-(1-q)\alpha_m\}^2 \\ &= (1-q)q\alpha_m^2,\end{aligned}$$

and

$$\begin{aligned}\sigma_{mat}^2 &= (1-q) \cdot (q\alpha_f)^2 + q \cdot \{-(1-q)\alpha_f\}^2 \\ &= (1-q)q\alpha_f^2.\end{aligned}$$

The sum of these variances is as follows:

$$\begin{aligned}\sigma_{pat}^2 + \sigma_{mat}^2 &= pq\alpha_m^2 + pq\alpha_f^2 \\ &= 2(1-q)q\{a + (2q-1)\delta\}^2 + 2(1-q)q\varepsilon^2 \\ &= \sigma_{a'}^2 + \sigma_{i'}^2.\end{aligned}$$

Thus, the total genetic variance can be partitioned as follows:

$$\sigma_g^2 = \sigma_{pat}^2 + \sigma_{mat}^2 + \sigma_{d'}^2.$$

### Statistical model

Two statistical models of GBLUP-I based on genotypic values (GBLUP-I1) and gametic values (GBLUP-I2) are proposed here.

First, GBLUP-I1 is defined as follows:

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

where  $\mathbf{y}$  is the vector of the phenotypes;  $\boldsymbol{\beta}$  is the vector of the fixed effects;  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{i}$  are the vectors of  $\alpha$ ,  $\delta$ , and  $\varepsilon$  terms, respectively;  $\mathbf{X}$ ,  $\mathbf{Z}_a$ ,  $\mathbf{Z}_d$ , and  $\mathbf{Z}_i$  are incidence matrices linking the phenotypes to  $\boldsymbol{\beta}$ ,  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{i}$ , respectively; and  $\mathbf{e}$  is the vector of errors. The variances of  $\mathbf{a}$ ,  $\mathbf{d}$ , and  $\mathbf{i}$  are as follows:

$$\text{Var}(\mathbf{a}) = \mathbf{G}_a\sigma_{a'}^2,$$

$$\text{Var}(\mathbf{d}) = \mathbf{G}_d\sigma_{d'}^2,$$

and

$$\text{Var}(\mathbf{i}) = \mathbf{G}_i\sigma_{i'}^2,$$

**Table 1 Average effects of paternal and maternal alleles at a QTL with imprinting**

Gamete type	Allele	Values and frequencies of genotypes produced				Mean value of genotypes produced	Average allele effect
		$A_1A_1$	$A_1A_2$	$A_2A_1$	$A_2A_2$		
		$a$	$\delta + \varepsilon$	$\delta - \varepsilon$	$-a$		
Sire	$A_1$	$1-q$	$q$			$(1-q)a + q(\delta + \varepsilon)$	$q\{a + (2q-1)\delta + \varepsilon\} = qa_m$
	$A_2$			$1-q$	$q$	$-qa + (1-q)(\delta - \varepsilon)$	$-(1-q)\{a + (2q-1)\delta + \varepsilon\} = -(1-q)a_m$
Dam	$A_1$	$1-q$		$q$		$(1-q)a + q(\delta - \varepsilon)$	$q\{a + (2q-1)\delta - \varepsilon\} = qa_f$
	$A_2$		$1-q$		$q$	$-qa + (1-q)(\delta + \varepsilon)$	$-(1-q)\{a + (2q-1)\delta - \varepsilon\} = -(1-q)a_f$

$a_m = a + (2q - 1)\delta + \varepsilon$ ;  $a_f = a + (2q - 1)\delta - \varepsilon$ ;  $a$  = genotypic value of  $A_1A_1$ ;  $\delta$  = mean of two heterozygotes;  $\varepsilon$  = difference between two heterozygotes;  $q$  = frequency of allele  $A_2$ .

where  $\mathbf{G}_a$ ,  $\mathbf{G}_d$ , and  $\mathbf{G}_i$  are the genomic relationship matrices relevant to  $\alpha$ ,  $\delta$ , and  $\varepsilon$  terms, respectively. These matrices describe the relationships among genotyped individuals and can be constructed by using the information from genome-wide SNPs. Let  $A_{1j}$  and  $A_{2j}$  be two alleles at the  $j^{\text{th}}$  SNP and  $q_j$  be the frequency of  $A_{2j}$ .  $\mathbf{G}_a$  and  $\mathbf{G}_d$  are the same as the genomic relationship matrices for breeding values and dominance deviations without imprinting. Thus,  $\mathbf{G}_a$  and  $\mathbf{G}_d$  can be calculated as described previously [21,22]:

$$\mathbf{G}_a = \frac{\mathbf{M}_a \mathbf{M}_a'}{\sum_j^{N_{\text{snp}}} 2q_j(1-q_j)},$$

and

$$\mathbf{G}_d = \frac{\mathbf{M}_d \mathbf{M}_d'}{\sum_j^{N_{\text{snp}}} \{2q_j(1-q_j)\}^2},$$

where  $\mathbf{M}_a$  and  $\mathbf{M}_d$  are  $n \times N_{\text{snp}}$  matrices ( $n$  is the number of genotyped individuals, and  $N_{\text{snp}}$  is the number of SNPs); the elements of  $\mathbf{M}_a$  and  $\mathbf{M}_d$  for the  $i^{\text{th}}$  individual at the  $j^{\text{th}}$  SNP are calculated as follows:

$$\mathbf{M}_a_{ij} = \begin{cases} 2q_j (A_1 A_1) \\ 2q_j - 1 (A_1 A_2 \text{ and } A_2 A_1) \\ 2q_j - 2 (A_2 A_2) \end{cases}$$

and

$$\mathbf{M}_d_{ij} = \begin{cases} -2q_j^2 (A_1 A_1) \\ 2q_j(1-q_j) (A_1 A_2 \text{ and } A_2 A_1) \\ -2(1-q_j)^2 (A_2 A_2) \end{cases}$$

Similarly,  $\mathbf{M}_i$  is assumed to be a  $n \times N_{\text{snp}}$  matrix, and the element of  $\mathbf{M}_i$  for the  $i^{\text{th}}$  individual at the  $j^{\text{th}}$  SNP can be calculated as follows:

$$\mathbf{M}_i_{ij} = \begin{cases} 0 (A_1 A_1) \\ 1 (A_1 A_2) \\ -1 (A_2 A_1) \\ 0 (A_2 A_2) \end{cases}$$

The elements of  $\mathbf{M}_a$ ,  $\mathbf{M}_d$ , and  $\mathbf{M}_i$  describe the coefficients of the  $\alpha$ ,  $\delta$ , and  $\varepsilon$  terms in Table 2, respectively. Therefore,  $\mathbf{i}$  and its variance can be derived as follows:

$$\mathbf{i} = \mathbf{M}_i \boldsymbol{\varepsilon},$$

where  $\boldsymbol{\varepsilon}$  is the  $N_{\text{snp}}$  dimensional vector of which the  $j^{\text{th}}$  element is  $\varepsilon_j$ . Thus, the variance of  $\mathbf{i}$  is calculated as follows:

$$\text{Var}(\mathbf{i}) = \mathbf{M}_i \mathbf{M}_i' \text{Var}(\boldsymbol{\varepsilon}),$$

$$\sigma_i^2 = \sum_j^{N_{\text{snp}}} 2q_j(1-q_j) \text{Var}(\varepsilon).$$

Consequently,  $\mathbf{G}_i$  can be calculated using  $\mathbf{M}_i$ :

$$\mathbf{G}_i = \frac{\mathbf{M}_i \mathbf{M}_i'}{\sum_j^{N_{\text{snp}}} 2q_j(1-q_j)}.$$

In general, GBLUP includes only breeding values. The statistical model of GBLUP is as follows:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_a \mathbf{a} + \mathbf{e}.$$

Therefore, without imprinting and dominance, the GBLUP model is the same as GBLUP-I1.

Second, GBLUP-I2 is defined as:

$$\mathbf{y} = \mathbf{X}\boldsymbol{\beta} + \mathbf{Z}_{\text{pat}} \mathbf{p}_{\text{at}} + \mathbf{Z}_{\text{mat}} \mathbf{m}_{\text{at}} + \mathbf{Z}_d \mathbf{d} + \mathbf{e},$$

where  $\mathbf{p}_{\text{at}}$  and  $\mathbf{m}_{\text{at}}$  are the vectors of paternal and maternal gametic effects, respectively; and  $\mathbf{Z}_{\text{pat}}$  and  $\mathbf{Z}_{\text{mat}}$  are incidence matrices linking phenotypes to  $\mathbf{p}_{\text{at}}$  and  $\mathbf{m}_{\text{at}}$ , respectively. The variances of  $\mathbf{p}_{\text{at}}$  and  $\mathbf{m}_{\text{at}}$  are as follows:

$$\text{Var}(\mathbf{p}_{\text{at}}) = \mathbf{G}_{\text{pat}} \sigma_{p_{\text{at}}}^2,$$

and

$$\text{Var}(\mathbf{m}_{\text{at}}) = \mathbf{G}_{\text{mat}} \sigma_{m_{\text{at}}}^2,$$

where  $\mathbf{G}_{\text{pat}}$  and  $\mathbf{G}_{\text{mat}}$  are the genomic relationship matrices of the paternal and maternal gametes, respectively. Let  $\mathbf{M}_{\text{pat}}$  and  $\mathbf{M}_{\text{mat}}$  be the  $n \times N_{\text{snp}}$  matrices that specify the coefficients of  $a_m$  and  $a_f$  in Table 1; then, the elements of  $\mathbf{M}_{\text{pat}}$  and  $\mathbf{M}_{\text{mat}}$  for the  $i^{\text{th}}$  individual at the  $j^{\text{th}}$  SNP are calculated as follows:

$$\mathbf{M}_{\text{pat}}_{ij} = \begin{cases} q_j & (A_1) \\ -(1-q_j) & (A_2) \end{cases},$$

and

$$\mathbf{M}_{\text{mat}}_{ij} = \begin{cases} q_j & (A_1) \\ -(1-q_j) & (A_2) \end{cases}.$$

Therefore,  $\mathbf{p}_{\text{at}}$  and  $\mathbf{m}_{\text{at}}$  are as follows:

$$\mathbf{p}_{\text{at}} = \mathbf{M}_{\text{pat}} \boldsymbol{\alpha}_m$$

and

$$\mathbf{m}_{\text{at}} = \mathbf{M}_{\text{mat}} \boldsymbol{\alpha}_f,$$

**Table 2 Genotypic values in the two-allele model**

	$A_1A_1$	$A_1A_2$	$A_2A_1$	$A_2A_2$
Genotypic value	$a$	$\delta + \varepsilon$	$\delta - \varepsilon$	$-a$
Deviation from population mean	$2qa - 2(1-q)q\delta$	$(2q-1)a + \{1-2(1-q)q\}\delta + \varepsilon$	$(2q-1)a + \{1-2(1-q)q\}q\delta - \varepsilon$	$-2(1-q)a - 2(1-q)q\delta$
$a$ term	$2qa$	$(2q-1)a$	$(2q-1)a$	$-2(1-q)a$
$\delta$ term	$-2q^2\delta$	$2(1-q)q\delta$	$2(1-q)q\delta$	$-2(1-q)^2\delta$
$\varepsilon$ term	0	$\varepsilon$	$-\varepsilon$	0

$a = a + (2q-1)\delta$ ;  $a$  = genotypic value of  $A_1A_1$ ;  $\delta$  = mean of two heterozygotes;  $\varepsilon$  = difference between two heterozygotes;  $q$  = frequency of allele  $A_2$ .

where  $\alpha_m$  and  $\alpha_f$  are the  $N_{snp}$ -dimensional vectors of  $\alpha_m$  and  $\alpha_f$  respectively. The variance of  $\mathbf{p}_{at}$  is equal to:

$$\text{Var}(\mathbf{p}_{at}) = \mathbf{M}_{\mathbf{p}_{at}} \mathbf{M}_{\mathbf{p}_{at}}' \text{Var}(\alpha_m).$$

The variance of the paternal gametic effect ( $\sigma_{pat}^2$ ) is the sum of the variances of  $\alpha_m$  at all SNPs as follows:

$$\sigma_{pat}^2 = \sum_j^{N_{snp}} q_j (1-q_j) \text{Var}(\alpha_m).$$

From this equation,  $\text{Var}(\mathbf{p}_{at})$  can be rewritten as follows:

$$\text{Var}(\mathbf{p}_{at}) = \frac{\mathbf{M}_{\mathbf{p}_{at}} \mathbf{M}_{\mathbf{p}_{at}}'}{\sum_j^{N_{snp}} q_j (1-q_j)} \sigma_{pat}^2.$$

Consequently,

$$\mathbf{G}_{\mathbf{p}_{at}} = \frac{\mathbf{M}_{\mathbf{p}_{at}} \mathbf{M}_{\mathbf{p}_{at}}'}{\sum_j^{N_{snp}} q_j (1-q_j)}.$$

Similarly,

$$\mathbf{G}_{\mathbf{m}_{at}} = \frac{\mathbf{M}_{\mathbf{m}_{at}} \mathbf{M}_{\mathbf{m}_{at}}'}{\sum_j^{N_{snp}} q_j (1-q_j)}.$$

### Stochastic simulation

A historical population was simulated to establish mutation-drift equilibrium. The simulated genome comprised 10 chromosomes, each 1 Morgan long, containing 100 000 randomly spaced SNPs and 1000 biallelic quantitative trait loci (QTL). In the first generation of the historical population, the initial allele frequencies of all SNPs and QTL were assumed to be 0.5. A recurrent mutation process was applied with a mutation rate for SNPs and QTL of  $1.0 \times 10^{-4}$  per locus per generation. Recombinations were sampled from a Poisson distribution with a mean of 1 per Morgan and then randomly placed along the chromosome. The historical population evolved over 20 000 generations of random selection and random

mating, with a population size of 500 (250 males and 250 females) to reach mutation-drift equilibrium [23].

After 20 000 historical generations, the base population (G0) was generated. In G0, the population size decreased to 300 (150 males and 150 females). 10 000 markers and 200 QTL were randomly selected from the segregating SNPs and QTL with minor allele frequencies greater than 0.05. Therefore,  $N_{snp}$  was equal to 10 000. Let  $Q_1$  and  $Q_2$  be two alleles at each QTL. The genotypic values of  $Q_1Q_1$ ,  $Q_1Q_2$ ,  $Q_2Q_1$  and  $Q_2Q_2$ , are given by  $a$ ,  $d_1$ ,  $d_2$  and  $-a$ , respectively. The value of  $a$  was drawn from a gamma distribution with a shape parameter of 0.42 and its sign was drawn at random with equal chance. For QTL with imprinting, the values of  $d_1$  and  $d_2$  were determined as the product of  $a$  and the degree of imprinting ( $r$ ). Let  $N_m$  and  $N_f$  be the number of QTL that are silencing the paternal alleles and maternal alleles. The total number of QTL with imprinting ( $N_i$ ) was 60 ( $N_m + N_f = N_i$ ), which were randomly chosen from the 200 QTL. The total genetic effect ( $g_j$ ) of the  $j^{th}$  animal was calculated by summing all QTL genotypic values, and its variance ( $\sigma_g^2$ ) was calculated from the variance of the genotypic deviations:

$$\begin{aligned} \sigma_g^2 = & \sum_{j=1}^{N_{QTL}} (1-q_j)^2 \{2q_j a_j - 2(1-q_j)q_j \delta_j\}^2 \\ & + \sum_{j=1}^{N_{QTL}} q_j (1-q_j) \left[ (2q_j - 1)a_j + \{1 - 2q_j(1-q_j)\} \delta_j + \varepsilon_j \right]^2 \\ & + \sum_{j=1}^{N_{QTL}} q_j (1-q_j) \left[ (2q_j - 1)a_j + \{1 - 2q_j(1-q_j)\} \delta_j - \varepsilon_j \right]^2 \\ & + \sum_{j=1}^{N_{QTL}} q_j^2 \{ -2(1-q_j)a_j - 2(1-q_j)q_j \delta_j \}^2, \end{aligned}$$

where  $N_{QTL}$  is the number of QTL. To obtain phenotypic values, an environmental effect was added to the true genetic value, which was sampled from the normal distribution,  $N(0, (1-H^2)\sigma_g^2/H^2)$ , where  $H^2$  is broad-sense heritability; narrow-sense heritability was set to 0.3. The phenotypic variance was finally standardized to be equal to 1.



After G0, the subsequent five generations (G1 to G5) were generated. In G1 to G5, 30 males were selected by BLUP on the basis of estimated breeding values and randomly mated to 150 dams to produce 300 offspring (150 males and 150 females). The reference population with both phenotypes and genotypes comprised 1200 individuals from G1 to G4, and the test population with only genotypes comprised 300 individuals from G5.

The range of  $d_1$  and  $d_2$ , the number of QTL with imprinting ( $N_i$ ), and  $N_{snp}$  were varied to investigate their effects on the performance of GBLUP-I. In the base simulation scenario,  $\tau = 1.0$ ,  $N_i = 60$ , ( $N_m$ ,  $N_f$ ) = (0, 60),  $N_{snp} = 10\ 000$ , and paternal and maternal alleles were known. In this scenario, only maternal alleles were silenced. Six alternative scenarios were simulated in addition to the base scenario. In scenario 1,  $\tau = 0.5, 0.75$ , and 1.0 to meet the condition that  $-a \leq d_1$ ,  $d_2 \leq a$ . In scenario 2,  $N_i = 20, 60$ , and 100. In scenario 3, ( $N_m$ ,  $N_f$ ) = (0, 60), (15, 45), and (30, 30). In scenario 4,  $H^2 = 0.1, 0.3$ , and 0.5. In scenario 5,  $N_{snp} = 2000, 10\ 000$ , and 50 000. In scenario 6, the paternal and maternal alleles were assumed to be unknown. Parameter settings are outlined in Table 3. Twenty replicates were simulated for each scenario.

#### Outline of the analysis

In the base scenario, the paternal and maternal alleles were assumed to be known. However, such information is unknown when using real data, because only genotypes are available. In scenario 6, the maternal and paternal origins of specific alleles (phase) were predicted using genotype and pedigree information processed by AlphaImpute software [24]. The phasing accuracy was measured as the correlation between true and predicted alleles by origin.

Here, we estimated variance components and genetic values using GBLUP and two types of GBLUP-I. Variance components were estimated by average information restricted maximum likelihood (AI-REML) [25]. The reference population dataset was used to predict the genetic effects of the genotyped individuals in the test

population. The accuracy of the estimated total genetic value ( $\rho$ ) was assessed as the correlation between estimated and true values. The regression coefficients of total genetic value on its estimate ( $b$ ) was calculated to assess unbiasedness.

#### Results

Tables 4 and 5 show the estimates of variance components and the predictive abilities of total genetic values with varying values of  $\tau$  and  $N_i$  in scenarios 1 and 2. Total genetic variance was underestimated by GBLUP when the degree of imprinting was high. With GBLUP, the estimated total genetic variances were equal to 97.6%, 91.3%, and 82.1% of true variances for  $\tau$  of 0.5, 0.75, and 1.0, respectively, and 99.3%, 82.1%, and 78.2% of true variances for  $N_i$  of 20, 50, and 100, respectively. The estimated total genetic variances by GBLUP-I1 and GBLUP-I2 were almost the same as the true variances regardless of the degree of imprinting.

The prediction accuracies,  $\rho$  obtained with GBLUP-I1 exceeded those obtained with GBLUP by 1.4%, 3.1%, and 7.8% for  $\tau$  of 0.5, 0.75, and 1.0, respectively, and by 0.2%, 7.8%, and 8.2% for  $N_i$  of 20, 50, and 100, respectively. Compared to GBLUP-I1, the  $\rho$  values obtained with GBLUP-I2 were more affected by the degree of imprinting. When  $N_i$  was equal to 60 and 100, the  $\rho$  values obtained with GBLUP-I2 exceeded those obtained with GBLUP by 6.8% and 11.0%; while, when  $N_i$  was equal to 20,  $\rho$  was smaller with GBLUP-I2 than with GBLUP. For all values of  $\tau$  and  $N_i$  the  $b$  values obtained with GBLUP-I1 and GBLUP-I2 were closer to 1 than with GBLUP. In scenario 3, the predictive abilities of GBLUP-I1 were not affected by the values of  $N_m$  and  $N_f$  whereas the  $\rho$  values with GBLUP-I2 decreased as the difference between  $N_m$  and  $N_f$  decreased (Table 6).

In scenario 4, for all values of  $H^2$  the estimated variance components obtained with GBLUP-I1 and GBLUP-I2 were close to the true values (Table 7). The performance of GBLUP-I1 and GBLUP-I2 increased with increasing values of  $H^2$ . With  $H^2$  of 0.1, 0.3, and 0.5, the  $\rho$  values obtained with GBLUP exceeded those obtained with

**Table 3 Parameters for different scenarios**

Parameter	Scenario						
	Base	1	2	3	4	5	6
$\tau$	1.0	0.5, 0.75, 1.0	1.0	1.0	1.0	1.0	1.0
$N_i$	60	60	20, 60, 100	60	60	60	60
( $N_m$ , $N_f$ )	(0, 60)	(0, 60)	(0, 60)	(0, 60)	(0, 60), (15, 45), (30, 30)	(0, 60)	(0, 60)
$H^2$	0.3	0.3	0.3	0.1, 0.3, 0.5	0.3	0.3	0.3
$N_{snp}$	10 000	10 000	10 000	10 000	10 000	2000, 10 000, 50 000	10 000
Paternal and maternal alleles	Known	Known	Known	Known	Known	Known	Known, predicted

Six alternative scenarios were simulated in addition to the base scenario:  $\tau$  = degree of imprinting;  $N_i$  = number of QTL with imprinting;  $N_m$  and  $N_f$  = numbers of QTL silencing paternal and maternal alleles;  $H^2$  = broad-sense heritability;  $N_{snp}$  = number of SNPs.

**Table 4 Variance component estimates and predictive abilities with varying degrees of imprinting ( $\tau$ ) in scenario 1**

$\tau$	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
0.5	True value	0.293	0.698	-	-
	GBLUP	0.286	0.692	0.626	1.026
	GBLUP-I1	0.300	0.679	0.635	1.010
	GBLUP-I2	0.294	0.679	0.570	0.996
0.75	True value	0.290	0.705	-	-
	GBLUP	0.265	0.716	0.581	1.011
	GBLUP-I1	0.295	0.681	0.599	1.000
	GBLUP-I2	0.300	0.680	0.570	0.989
1.0	True value	0.291	0.701	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.299	0.681	0.570	0.990
	GBLUP-I2	0.297	0.681	0.565	0.995

Values are the mean of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

GBLUP-I1 by 5.7%, 7.8%, and 9.2% and those obtained with GBLUP-I2 by 4.1%, 6.8%, and 7.1%. In scenario 5, the predictive abilities of GBLUP, GBLUP-I1, and GBLUP-I2 decreased when  $N_{\text{snp}}$  decreased from 10 000 to 2000 whereas those were unaltered when  $N_{\text{snp}}$  increased from 10 000 to 50 000 (Table 8).

In scenario 6, the phasing accuracy was equal to 0.979. Prediction accuracies with GBLUP-I1 and GBLUP-I2 were 1.7% and 1.2% lower when paternal and maternal

alleles were predicted than when paternal and maternal alleles were known (Table 9).

## Discussion

### Performance of GBLUP-I

We present a new GBLUP method that includes imprinting effects for the prediction of total genetic value. For all scenarios, the performance of GBLUP-I1 to estimate variance components was always better than that of GBLUP. Prediction accuracies with GBLUP-I1 and GBLUP-I2 increased with increasing degree of imprinting and broad-sense heritability. Prediction accuracies with GBLUP-I2 were strongly affected by the degree of imprinting (Tables 4 and 5) and the difference between the values of  $N_m$  and  $N_f$  (Table 6).

Method GBLUP-I2 assumes that paternal and maternal gametic effects are independent. However, when there is no imprinting, sire and dam are genetically correlated, and thus paternal and maternal gametic effects are not independent [26] and the accuracy by GBLUP-I2 should be reduced. The reduction of accuracy by GBLUP-I2 would be small with a high degree of imprinting because of the low correlation between paternal and maternal gametic effects. Thus, the performance of GBLUP-I2 increases as the degree of the imprinting and the difference in genetic values between paternal and maternal gametes increase. Meanwhile, when the degree of imprinting is low and there is little difference in the genetic values between paternal and maternal gametes, GBLUP-I1 is preferred for genomic evaluation.

In a previous study with bovine data, when the number of SNPs was greater than 50 000, reliabilities of

**Table 5 Variance component estimates and predictive ability with varying numbers of QTL with imprinting ( $N_i$ ) in scenario 2**

$N_i$	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
20	True value	0.289	0.705	-	-
	GBLUP	0.287	0.696	0.659	1.078
	GBLUP-I1	0.291	0.689	0.660	1.076
	GBLUP-I2	0.286	0.688	0.596	1.071
60	True value	0.291	0.701	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.299	0.681	0.570	0.990
	GBLUP-I2	0.297	0.681	0.565	0.995
100	True value	0.293	0.698	-	-
	GBLUP	0.229	0.756	0.518	0.975
	GBLUP-I1	0.298	0.683	0.564	0.982
	GBLUP-I2	0.295	0.664	0.575	0.999

Values are the mean of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

**Table 6 Variance component estimates and predictive ability with varying numbers of QTL silencing paternal and maternal alleles ( $N_m$  and  $N_f$ ) in scenario 3**

$N_m$	$N_f$	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
0	60	True value	0.291	0.701	-	-
		GBLUP	0.239	0.742	0.529	0.982
		GBLUP-I1	0.299	0.681	0.570	0.990
		GBLUP-I2	0.297	0.681	0.565	0.995
15	45	True value	0.295	0.698	-	-
		GBLUP	0.234	0.739	0.519	0.981
		GBLUP-I1	0.296	0.686	0.569	1.002
		GBLUP-I2	0.298	0.688	0.549	0.998
30	30	True value	0.290	0.704	-	-
		GBLUP	0.234	0.739	0.519	0.981
		GBLUP-I1	0.296	0.686	0.573	1.009
		GBLUP-I2	0.295	0.684	0.538	0.998

Values are the mean of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.



**Table 7** Variance component estimates and predictive ability with varying broad-sense heritability ( $H^2$ ) in scenario 4

$H^2$	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
0.1	True value	0.096	0.902	-	-
	GBLUP	0.078	0.920	0.367	1.021
	GBLUP-I1	0.094	0.885	0.388	0.990
	GBLUP-I2	0.090	0.886	0.382	0.986
0.3	True value	0.291	0.701	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.299	0.681	0.570	0.990
	GBLUP-I2	0.297	0.681	0.565	0.995
0.5	True value	0.493	0.500	-	-
	GBLUP	0.403	0.597	0.608	1.010
	GBLUP-I1	0.499	0.492	0.664	1.001
	GBLUP-I2	0.498	0.492	0.651	1.000

Values are the mean of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

genomic evaluations remained almost unaltered as the number of SNPs increased [27]. In this study, with a  $N_{snp}$  of 10 000, the average distance between neighboring SNPs was 0.1 cM, which is similar to the distance between SNPs in a bovine dataset that includes 50 000 SNPs. In scenario 5, when  $N_{snp}$  was greater than 10 000, the performances of GBLUP and both GBLUP-I were not affected by various values of  $N_{snp}$ . This suggests that high-density and costly chips with more (777 000) SNPs may not be necessary for genomic evaluation, even when imprinting effects exist.

**Table 8** Variance component estimates and predictive ability with varying numbers of SNPs ( $N_{snp}$ ) in scenario 5

$N_{snp}$	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
2000	True value	0.290	0.694	-	-
	GBLUP	0.201	0.770	0.481	0.939
	GBLUP-I1	0.263	0.727	0.521	0.946
	GBLUP-I2	0.256	0.733	0.514	0.948
10 000	True value	0.291	0.701	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.299	0.681	0.570	0.990
	GBLUP-I2	0.297	0.681	0.565	0.995
50 000	True value	0.292	0.693	-	-
	GBLUP	0.242	0.725	0.536	1.072
	GBLUP-I1	0.296	0.687	0.572	1.070
	GBLUP-I2	0.294	0.694	0.566	1.048

Values are the mean of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

**Table 9** Accuracies of estimated genetic values with predicted paternal and maternal alleles in scenario 6

Phasing accuracy	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
1.0	True value	0.291	0.701	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.299	0.681	0.570	0.990
	GBLUP-I2	0.297	0.681	0.565	0.995
0.979	True value	0.289	0.705	-	-
	GBLUP	0.239	0.742	0.529	0.982
	GBLUP-I1	0.295	0.690	0.560	0.984
	GBLUP-I2	0.294	0.689	0.558	0.991

Values are the mean (standard error) of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance; predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

In scenario 6, prediction accuracies obtained with GBLUP-I1 and GBLUP-I2 were higher than those obtained with GBLUP when paternal and maternal alleles were predicted. Thus, both GBLUP-I methods can be applied to real livestock data. The phasing accuracy was improved by increasing sample size [28], number of SNPs [29], and number of high-density genotyped relatives of the individuals to be imputed [24,30], which suggests that the performance of GBLUP-I can be further improved in real livestock data.

**Degree of imprinting and number of QTL with imprinting**  
GBLUP-I1 and GBLUP-I2 could accurately capture the total genetic variance, whereas GBLUP underestimated the total genetic variance. The difference in estimated total genetic variance between GBLUP-I and GBLUP is caused by the imprinting effect. Here, we calculated imprinting variance as the difference in the estimated total genetic variance between GBLUP-I1 and GBLUP. When  $\tau$  varied from 0.5 to 1.0 and  $N_i$  from 20 to 100, imprinting variances were equal to 1.4% to 6.0% and 0.4% to 6.9% of the phenotypic variances (1.0), respectively.

de Vries et al. [31] were the first to estimate imprinting variance in livestock and found that approximately 5% and 4% of the phenotypic variance in back fat thickness and growth rate, respectively, were due to imprinting. More recently, imprinting variances were found to range from 5 to 19% of the total genetic variance for 19 pig performance traits [13] and, on average, to be equal to 28% of the total genetic variance for ultrasonic measurements of body composition in Australian beef cattle [26]. The degree of imprinting reported in our study is similar to those reported in the literature [13,26,31].

#### Effects of QTL parameters

Setting QTL parameters may affect the accuracy of genomic predictions. We investigated the effects of the

number of QTL, the distribution of their effects, and their location. The values of  $N_{QTL}$  ranged from 50 ( $N_i = 15$ ) to 1000 ( $N_i = 300$ ) and QTL were evenly spaced throughout the genome. The value of  $a$  was drawn from a normal distribution. In these conditions, the accuracies obtained by GBLUP, GBLUP-I1, and GBLUP-I2 were the almost the same as in the base scenario (Table 10).

#### Significance of genetic effects in GBLUP-I

This study partitioned the total genetic value into three estimated genetic effects ( $\alpha$ ,  $\delta$ , and  $\varepsilon$  terms) in GBLUP-I1. However, there is no biological meaning for these genetic effects. In order to estimate a breeding value and a dominance deviation, the genetic values should be defined by sex, as presented by Spencer [19]. In such a model, the number of variance components would be doubled and the covariance between the breeding value and dominance deviation would not be equal to 0. These factors would collectively reduce the accuracy of genetic evaluations.

#### Practical use of GBLUP-I

When a dominance effect exists, assortative mating or mate allocation can boost the field performance of livestock [32,33]. Similarly, when an imprinting effect exists, the performance of livestock can be improved by optimizing matings, because the genotypic values of  $A_1A_2$  and  $A_2A_1$  can be distinguished and evaluated accurately. Let  $pr_{ij}(A_1A_1)$   $pr_{ij}(A_1A_2)$   $pr_{ij}(A_2A_1)$  and  $pr_{ij}(A_2A_2)$  be

the probabilities of the genotypes  $A_1A_1$ ,  $A_1A_2$ ,  $A_2A_1$ , and  $A_2A_2$  for the  $i^{th}$  offspring of future matings and the  $j^{th}$  marker. In GBLUP-I1, the elements of coefficient matrices for these offspring (i.e.,  $\mathbf{M}_a$ ,  $\mathbf{M}_d$ , and  $\mathbf{M}_i$ ) can be calculated from the products of coefficients and the genotype probabilities. For example, the element of  $\mathbf{M}_i$  for offspring is as follows:

$$\mathbf{M}_{i,j} = \begin{cases} 0 (A_1A_1) \\ 1 \times pr_{ij}(A_1A_2) (A_1A_2) \\ -1 \times pr_{ij}(A_2A_1) (A_2A_1) \\ 0 (A_2A_2) \end{cases}$$

Likewise, in GBLUP-I2, the elements of  $\mathbf{M}_{pat}$  and  $\mathbf{M}_{mat}$  for the offspring of future matings can be calculated. Thus, the total genetic effects for the offspring of future matings can be predicted and maximized by using an optimum mating plan.

#### Conclusions

This study proposed two GBLUP methods i.e., GBLUP-I1 and GBLUP-I2, which include imprinting effects at the genotypic and gametic levels, respectively. The GBLUP-I1 and GBLUP-I2 methods accurately estimated the variance components and improved unbiasedness regardless of parameter settings. The accuracies of estimated total genetic values in GBLUP-I1 and GBLUP-I2 increased with increasing degree of imprinting and broad-sense heritability. Compared to GBLUP, the accuracies of estimated total

**Table 10 Accuracies of estimated genetic values with varying numbers of QTL, distributions of homozygous genotypic value, and locations of QTL**

Number of QTL	Distribution of homozygous genotypic value ( $a$ )	QTL location	Method	$\sigma_g^2$	$\sigma_e^2$	$\rho$	$b$
50	Gamma	Random	GBLUP	0.238	0.740	0.528	0.983
			GBLUP-I1	0.295	0.683	0.565	0.989
			GBLUP-I2	0.296	0.685	0.564	0.992
200	Gamma	Random	GBLUP	0.239	0.742	0.529	0.982
			GBLUP-I1	0.299	0.681	0.570	0.990
			GBLUP-I2	0.297	0.681	0.565	0.995
1000	Gamma	Random	GBLUP	0.235	0.745	0.527	0.979
			GBLUP-I1	0.297	0.686	0.568	0.994
			GBLUP-I2	0.294	0.684	0.563	0.993
200	Normal	Random	GBLUP	0.239	0.741	0.529	0.980
			GBLUP-I1	0.294	0.681	0.566	0.991
			GBLUP-I2	0.295	0.683	0.560	0.994
200	Gamma	Evenly spaced	GBLUP	0.245	0.738	0.531	0.989
			GBLUP-I1	0.295	0.688	0.572	0.997
			GBLUP-I2	0.294	0.689	0.567	0.999

Values are the means of 20 replicates; variance components for each source of genetic variation:  $\sigma_g^2$  = total genetic variance;  $\sigma_e^2$  = residual variance. Predictive abilities:  $\rho$  = accuracy of estimated total genetic value;  $b$  = regression coefficient of total genetic value on its estimate.

genetic values obtained with GBLUP-I1 were always higher. Thus, in general, GBLUP-I1 should be applied for genetic evaluation. However, GBLUP-I2 is preferred when the imprinting effect is large and the genetic effects differ substantially between paternal and maternal gametes. After predicting the total genetic value by both GBLUP-I methods, assortative mating or mate allocation could be used to boost the field performance of livestock.

#### Competing interests

The authors declare that they have no competing interests.

#### Authors' contributions

MN developed the two GBLUP models that include the imprinting effect, wrote all the computer programs, and drafted the manuscript. MS conceived and set up the study, and helped with writing the manuscript. Both authors have read and approved the final manuscript.

#### Acknowledgements

The authors thank John Hickey for kindly providing the Alphalpmpute software.

Received: 13 November 2014 Accepted: 14 January 2015

Published online: 19 April 2015

#### References

- Reik W, Walter J. Genomic imprinting: parental influence on the genome. *Nat Rev Genet*. 2001;2:21–32.
- O'Neill MJ, Ingram RS, Vrana PB, Tilghman SM. Allelic expression of IGF2 in marsupials and birds. *Dev Genes Evol*. 2000;210:18–20.
- Georges M, Charlier C, Cockett N. The callipyge locus: evidence for the trans interaction of reciprocally imprinted genes. *Trends Genet*. 2003;19:248–52.
- Sakatani T, Wei M, Katoh M, Okita C, Wada D, Mitsuya K, et al. Epigenetic heterogeneity at imprinted loci in normal populations. *Biochem Biophys Res Commun*. 2001;283:1124–30.
- Morison IM, Ramsay JP, Spencer HG. A census of mammalian imprinting. *Trends Genet*. 2005;21:457–65.
- Imumori IG, Kim EH, Lee YM, de Koning DJ, van Arendonk JAM, Donato MD, et al. Genome scan for parent-of-origin QTL effects on bovine growth and carcass traits. *Front Genet*. 2011;2:44.
- de Koning DJ, Rattink AP, Harlizius B, Groenen MAM, Brascamp EW, van Arendonk JAM. Detection and characterization of quantitative trait loci for growth and reproduction traits in pigs. *Livest Prod Sci*. 2001;72:185–98.
- Quintanilla R, Milan D, Bidanel JP. A further look at quantitative trait loci affecting growth and fatness in a cross between Meishan and Large White pig populations. *Genet Sel Evol*. 2002;34:193–210.
- Hirooka H, de Koning DJ, Harlizius B, van Arendonk JAM, Rattink AP, Groenen MA, et al. A whole-genome scan for quantitative trait loci affecting teat number in pigs. *J Anim Sci*. 2001;79:2320–6.
- Stella A, Stalder KJ, Saxton AM, Boettcher PJ. Estimation of variances for gametic effects on litter size in Yorkshire and Landrace swine. *J Anim Sci*. 2003;81:2171–8.
- Schaeffer LR, Kennedy BW, Gibson JP. The inverse of the gametic relationship matrix. *J Dairy Sci*. 1989;72:1266–72.
- Essl A, Voith K. Genomic imprinting effects on dairy- and fitness-related traits in cattle. *J Anim Breed Genet*. 2002;119:182–9.
- Neugebauer N, Luther H, Reinsch N. Parent-of-origin effects cause genetic variation in pig performance traits. *Animal*. 2010;4:672–81.
- Neugebauer N, Rader I, Schild HJ, Zimmer D, Reinsch N. Evidence for parent-of-origin effects on genetic variability of beef traits. *J Anim Sci*. 2010;88:523–32.
- Dalton R. No bull: genes for better milk. *Nature*. 2009;457:369.
- Lund M, de Ross S, de Vries A, Druet T, Ducrocq V, Fritz S, et al. A common reference population from four European Holstein populations increases reliability of genomic predictions. *Genet Sel Evol*. 2011;43:43.
- Mc Hugh N, Meuwissen TH, Cromie AR, Sonesson AK. Use of female information in dairy cattle genomic breeding programs. *J Dairy Sci*. 2011;94:4109–18.
- Wiggans GR, VanRaden PM, Cooper TA. The genomic evaluation system in the United States: past, present, future. *J Dairy Sci*. 2011;94:3202–11.
- Spencer HG. The correlation between relatives on the supposition of genomic imprinting. *Genetics*. 2002;161:411–7.
- Falconer DS, Mackay TFC. *Introduction to Quantitative Genetics*. Addison Wesley Longman: Essex; 1996.
- VanRaden PM. Efficient methods to compute genomic predictions. *J Dairy Sci*. 2008;91:4414–23.
- Nishio M, Satoh M. Including dominance effects in the genomic BLUP method for genomic evaluation. *PLoS One*. 2014;9:e85792.
- Nishio M, Satoh M. Parameters affecting genome simulation for evaluating genomic selection method. *Anim Sci J*. 2014;85:879–87.
- Hickey JM, Kinghorn BP, Tier B, van der Werf JH, Cleveland MA. A phasing and imputation method for pedigreed populations that results in a single-stage genomic evaluation method. *Genet Sel Evol*. 2012;44:9.
- Johnson DL, Thompson R. Restricted maximum likelihood estimation of variance components for univariate animal models using sparse matrix techniques and average information. *J Dairy Sci*. 1995;78:449–56.
- Tier B, Meyer K. On the analysis of parent-of-origin effects with examples from ultrasonic measures of carcass traits in Australian beef cattle. *J Anim Breed Genet*. 2012;129:359–68.
- VanRaden PM, O'Connell JR, Wiggans GR, Weigel KA. Genomic evaluations with many more genotypes. *Genet Sel Evol*. 2011;43:10.
- Huang L, Li Y, Singleton AB, Hardy JA, Abecasis G, Rosenberg NA, et al. Genotype-imputation accuracy across worldwide human populations. *Am J Hum Genet*. 2009;84:235–50.
- Zhang Z, Druet T. Marker imputation with low-density marker panels in Dutch Holstein cattle. *J Dairy Sci*. 2010;93:5487–94.
- Hickey JM, Crossa J, Babu R, de los Campos G. Factors affecting the accuracy of genotype imputation in populations from several maize breeding programs. *Crop Sci*. 2012;52:654–63.
- de Vries AG, Kerr R, Tier B, Long T, Meuwissen TH. Gametic imprinting effects on rate and composition of pig growth. *Theor Appl Genet*. 1994;88:1037–42.
- Toro MA, Varona L. A note on mate allocation for dominance handling in genomic selection. *Genet Sel Evol*. 2010;42:33.
- Zeng J, Toosi A, Fernando RL, Dekkers JCM, Garrick DJ. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. *Genet Sel Evol*. 2013;45:11.

**Submit your next manuscript to BioMed Central and take full advantage of:**

- Convenient online submission
- Thorough peer review
- No space constraints or color figure charges
- Immediate publication on acceptance
- Inclusion in PubMed, CAS, Scopus and Google Scholar
- Research which is freely available for redistribution

Submit your manuscript at  
www.biomedcentral.com/submit

