Scientia Agricola
http://dx.doi.org/10.1590/1678-992X-2015-0479

Quantitative genetics theory for genomic selection and efficiency of genotypic value prediction in open-pollinated populations

José Marcelo Soriano Viana[1]*, Hans-Peter Piepho[2], Fabyano Fonseca e Silva[3]

[1]Federal University of Viçosa – Dept. of General Biology, Av. Peter Henry Rolfs, s/n – 36570-900 – Viçosa, MG – Brazil.
[2]University of Hohenheim/Institute of Crop Science – Biostatistics Unit, Fruwirthstrasse 23 – 70599 – Stuttgart – Germany.
[3]Federal University of Viçosa – Dept. of Animal Science.
*Corresponding author <jmsviana@ufv.br>

Edited by: Leonardo Oliveira Medici

Received December 10, 2015
Accepted March 21, 2016

ABSTRACT: Quantitative genetics theory for genomic selection has mainly focused on additive effects. This study presents quantitative genetics theory applied to genomic selection aiming to prove that prediction of genotypic value based on thousands of single nucleotide polymorphisms (SNPs) depends on linkage disequilibrium (LD) between markers and QTLs, assuming dominance and epistasis. Based on simulated data, we provided information on dominance and genotypic value prediction accuracy, assuming mass selection in an open-pollinated population, all quantitative trait loci (QTLs) of lower effect, and reduced sample size. We show that the predictor of dominance value is proportional to the square of the LD value and to the dominance deviation for each QTL that is in LD with each marker. The weighted (by the SNP frequencies) dominance value predictor has greater accuracy than the unweighted predictor. The linear × linear, linear × quadratic, quadratic × linear, and quadratic × quadratic SNP effects are proportional to the corresponding linear combinations of epistatic effects for QTLs and the LD values. LD between two markers with a common QTL causes a bias in the prediction of epistatic values. Compared to phenotypic selection, the efficiency of genomic selection for genotypic value prediction increases as trait heritability decreases. The degree of dominance did not affect the genotypic value prediction accuracy and the approach to maximum accuracy is asymptotic with increases in SNP density. The decrease in the sample size from 500 to 200 did not markedly reduce the genotypic value prediction accuracy.

Keywords: genome-wide selection, dominance value prediction, prediction accuracy

Introduction

The statistical problem of breeding value prediction when there are a very large number of markers and few observations has been addressed thorough various genomic selection models, most of which provide similar prediction accuracy levels (De Los Campos et al., 2013; Daetwyler et al., 2013). A current focus of genomic selection research is on fitting non-additive effects to predicting genotypic values (Wang et al., 2010; Zhao et al., 2013). For maize, where the main objective in commercial breeding programs is to develop hybrids, predicting genotypic value is as important as predicting breeding value aimed at exploring heterosis (Technow et al., 2012; Massman et al., 2013).

Toro and Varona (2010) analyzed simulated data and concluded that the inclusion of dominance effects increased the accuracy of breeding value prediction and made it possible to obtain an extra response to selection using mate allocation techniques. Wittenburg et al. (2011) included dominance and epistasis in simulated datasets. The analyses showed that only the inclusion of dominance effects improved breeding value prediction accuracy. Using simulated data and assuming one additive and three additive-dominance models, Wellmann and Bennewitz (2012) analyzed the prediction accuracy of additive, dominance, and genotypic values. The inclusion of dominance effects increased the accuracy of genotypic value prediction by approximately 17 % and the accuracy of breeding value prediction by 2 %. Denis and Bouvet (2013) assessed the prediction accuracy of additive and genotypic values fitting an additive-dominance model. Including the dominance effect improved the accuracy of genotypic value prediction in the clone population but not the accuracy of additive value prediction in the breeding population.

Quantitative genetics theory including non-additive effects does not seem to have been fully developed in the context of genomic prediction (Goddard, 2009; Gianola et al., 2009; Vitezica et al., 2013). Furthermore, efficient genome-wide prediction of the genotypic value of non-assessed single-crosses and pure lines, and of vegetative propagated plants in a recurrent breeding program, depends on the prediction accuracy of non-additive gene effects (ultimately, on the genotypic value prediction accuracy). Thus, we presented quantitative genetics theory applied to genomic selection aiming to prove that prediction of genotypic value based on thousands of single nucleotide polymorphisms (SNPs) depends on linkage disequilibrium (LD) between markers and quantitative trait loci (QTLs), assuming dominance and epistasis. Additionally, we provided information on dominance and genotypic values prediction accuracy based on simulated data, assuming mass selection in open-pollinated populations, QTLs of lower effect, and reduced sample size.

Materials and Methods

Theory

Linkage disequilibrium measure

It was assumed that the population was a Hardy-Weinberg equilibrium population (generation -1). It was further assumed that B and b are the alleles of a QTL (QTL 1) and that C and c are the alleles of an SNP (SNP 1). B is the allele that increases the trait expression and let b be the allele that decreases the trait expression. Assuming linkage, the probabilities of the gametes BC, Bc, bC, and bc in the gametic pool produced by the population are, respectively:

$$P_{BC}^{(-1)} = p_b p_c + \Delta_{bc}^{(-1)}$$

$$P_{Bc}^{(-1)} = p_b q_c - \Delta_{bc}^{(-1)}$$

$$P_{bC}^{(-1)} = q_b p_c - \Delta_{bc}^{(-1)}$$

$$P_{bc}^{(-1)} = q_b q_c + \Delta_{bc}^{(-1)}$$

where p is the frequency of the major allele (B or C), q = 1 – p is the frequency of the minor allele (b or c), and $\Delta_{bc}^{(-1)} = P_{BC}^{(-1)} P_{bc}^{(-1)} - P_{Bc}^{(-1)} P_{bC}^{(-1)}$ is the measure of LD (Kempthorne, 1957). The LD measure can also be expressed as $\Delta_{bc}^{(-1)} = r_{bc}^{(-1)} \sqrt{p_b q_b p_c q_c}$, where $r_{bc}^{(-1)}$ is the correlation between the values of the alleles at the two loci (one for B and C, and zero for b and c) in the gametic pool of generation -1 (Hill and Robertson, 1968).

Genetic model

It was assumed that E and e are the alleles of a second QTL (QTL 2), where E and e increase and decrease the trait expression, respectively. The genotypic value of an individual for the two QTLs can be defined as (Kempthorne, 1954):

$$\begin{aligned} G_{ijmn} &= M + \left(\alpha_i + \alpha_j\right) + \left(\alpha_m + \alpha_n\right) + \delta_{ij} + \delta_{mn} + \left(\alpha_i\alpha_m + \alpha_i\alpha_n + \alpha_j\alpha_m + \alpha_j\alpha_n\right) \\ &\quad + \left(\alpha_i\delta_{mn} + \alpha_j\delta_{mn}\right) + \left(\delta_{ij}\alpha_m + \delta_{ij}\alpha_n\right) + \delta_{ij}\delta_{mn} \\ &= M + A_1 + A_2 + D_1 + D_2 + AA_{(12)} + AD_{(12)} + DA_{(12)} + DD_{(12)} \end{aligned}$$

where i and j are the alleles for the first QTL, m and n are the alleles for the second QTL, M is the population mean, α is the average effect of a gene (for brevity we are dropping subscripts in the definition of effects), δ is the dominance genetic value, αα is the additive × additive epistatic effect (due to a pair of non-allelic genes), αδ is the additive × dominance epistatic effect (due to a gene of a locus and the alleles of a second locus), δα is the dominance × additive epistatic effect, δδ is the epistatic effect of the type dominance × dominance (due to two pair of allelic genes), A is the additive genetic value, D is the dominance genetic value, AA is the additive × additive epistatic genetic value, AD is the additive × dominance epistatic value, DA is the dominance × additive epistatic value, and DD is the dominance × dominance epistatic value.

Assuming biallelic QTLs, $\alpha_B = q_b\alpha_{\mathrm{b}}$, $\alpha_b = -p_b\alpha_{\mathrm{b}}$, $\alpha_E = q_e\alpha_{\mathrm{e}}$, and $\alpha_e = -p_e\alpha_{\mathrm{e}}$, where $\alpha_{\mathrm{b}} = \alpha_B - \alpha_b = a_b + (q_b - p_b) d_b$ and $\alpha_{\mathrm{e}} = \alpha_E - \alpha_e = \mathrm{a}_e + (q_e - p_e) d_e$ are the average effects of gene substitution. The parameter a is the deviation between the genotypic value of the homozygote of higher expression and the mean of the genotypic values of the homozygotes (m), and the parameter d is the deviation between the genotypic value of the heterozygote and m (dominance deviation).

The parametric values of the regression coefficients in a whole-genome analysis

F and f are the alleles of a second SNP (SNP 2), which is in LD with QTL 2. Finally, assume (for simplicity) that QTL 1 and SNP 1 are in linkage equilibrium relative to QTL 2 and SNP 2. The parametric values of the regression coefficients in a whole-genome analysis are derived by regression analysis that relates the genotypic value (G) conditional on the SNP genotype to the number of copies of one allele of each SNP. The additive-dominance with epistasis model is:

$$G = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \beta_5 x_1 x_2 + \beta_6 x_1 x_2^2 + \beta_7 x_1^2 x_2 + \beta_8 x_1^2 x_2^2 + error$$

where x_1 and x_2 are the numbers of copies of an SNP allele (2, 1, or 0).

The model can be expressed as y (81 × 1) = X (81 × 9).β (9 × 1) + error vector (81 × 1), where y is the vector of genotypic values conditional on the SNP genotypes, X is the incidence matrix, and β is the parameter vector. Notice that assuming biallelic QTLs, there are 3 × 3 × 3 × 3 = 81 genotypes for QTLs and markers (for example, BbCCeeFf). Because the genotypes have different probabilities, we defined the matrix of genotype probabilities as P (81 × 81) = diagonal $\left\{f_{ij}.f_{kl}^{'}\right\}$, where f_{ij} is the probability of the individual with i and j copies of allele B of the QTL 1 and allele C of the SNP 1, and $f_{kl}^{'}$ is the probability of the individual with k and l copies of allele E of the QTL 2 and allele F of the SNP 2 (i, j, k, l = 2, 1, or 0).

The genotype probabilities relative to QTL 1 and SNP 1 in generation 0 are (for simplicity, the superscript (0) – for generation 0 – was omitted in all parameters that depend on the LD measure of generation −1):

$$f_{22} = p_b^2 p_c^2 + 2p_b p_c \Delta_{bc}^{(-1)} + \left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{21} = 2p_b^2 p_c q_c + 2p_b\left(q_c - p_c\right)\Delta_{bc}^{(-1)} - 2\left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{20} = p_b^2 q_c^2 - 2p_b q_c \Delta_{bc}^{(-1)} + \left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{12} = 2p_b q_b p_b^2 + 2\left(q_b - p_b\right)p_c \Delta_{bc}^{(-1)} - 2\left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{11} = f_{11g} + f_{11n} = 4p_b q_b p_c q_c + 2\left(q_b - p_b\right)\left(q_c - p_c\right)\Delta_{bc}^{(-1)} + 4\left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{10} = 2p_b q_b q_c^2 - 2\left(q_b - p_b\right)q_c \Delta_{bc}^{(-1)} - 2\left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{02} = q_b^2 p_c^2 - 2q_b p_c \Delta_{bc}^{(-1)} + \left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{01} = 2q_b^2 p_c q_c - 2q_b\left(q_c - p_c\right)\Delta_{bc}^{(-1)} - 2\left[\Delta_{bc}^{(-1)}\right]^2$$

$$f_{00} = q_b^2 q_c^2 + 2q_b q_c \Delta_{bc}^{(-1)} + \left[\Delta_{bc}^{(-1)}\right]^2$$

where the indices g and n identify the double heterozygotes in coupling and repulsion phases.

Thus, for the complete model or a reduced model, $\beta = (X'PX)^{-1}(X'Py)$ and R(.) = $\beta'(X'Py)$, where R(.) is the reduction in the total sum of squares due to fitting the model. Finally, after fitting the complete model and eight reduced models, and assuming the same restrictions of Kempthorne (1954), it can be demonstrated that:

$\beta_0 = M$ (fitting $G = \beta_0 + \varepsilon$, the reduced model 1)

$\beta_1 = \left[\dfrac{\Delta_{bc}^{(-1)}}{p_c q_c}\right]\alpha_b = \kappa_{bc}\alpha_b = \alpha_{SNP1}$ (fitting $G = \beta_0 + \beta_1 x_1 + \varepsilon$, the reduced model 2)

$\beta_2 = -\kappa_{bc}^2 d_b = -d_{SNP1}$ (fitting $G = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \varepsilon$, the reduced model 3)

$\beta_3 = \left[\dfrac{\Delta_{ef}^{(-1)}}{p_f q_f}\right]\alpha_e = \kappa_{ef}\alpha_e = \alpha_{SNP2}$
(fitting $G = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \varepsilon$, the reduced model 4)

$\beta_4 = -\kappa_{ef}^2 d_e = -d_{SNP2}$
(fitting $G = \beta_0 + \beta_1 x_1 + \beta_2 x_1^2 + \beta_3 x_2 + \beta_4 x_2^2 + \varepsilon$, the reduced model 5)

$\beta_5 = \kappa_{bc}\kappa_{ef}(aa)_{be} = (\alpha\alpha)_{SNP1,SNP2}$
(fitting $G = \beta_0 + \ldots + \beta_5 x_1 x_2 + \varepsilon$, the reduced model 6)

$\beta_6 = (1/2)\kappa_{bc}\kappa_{ef}^2(ad)_{be} = (\alpha\delta)_{SNP1,SNP2}$
(fitting $G = \beta_0 + \ldots + \beta_6 x_1 x_2^2 + \varepsilon$, the reduced model 7)

$\beta_7 = (1/2)\kappa_{bc}^2\kappa_{ef}(da)_{be} = (\delta\alpha)_{SNP1,SNP2}$
(fitting $G = \beta_0 + \ldots + \beta_7 x_1^2 x_2 + \varepsilon$, the reduced model 8)

$\beta_8 = (1/4)\kappa_{bc}^2\kappa_{ef}^2(dd)_{be} = (\delta\delta)_{SNP1,SNP2}$ (fitting the complete model)

where: $(aa)_{be} = \alpha_B\alpha_E - \alpha_B\alpha_e - \alpha_b\alpha_E + \alpha_b\alpha_e$

$$(ad)_{be} = \alpha_B\delta_{EE} - 2\alpha_B\delta_{Ee} + \alpha_B\delta_{ee} - \alpha_b\delta_{EE} + 2\alpha_b\delta_{Ee} - \alpha_b\delta_{ee}$$

$$(da)_{be} = \delta_{BB}\alpha_E - 2\delta_{Bb}\alpha_E + \delta_{bb}\alpha_E - \delta_{BB}\alpha_e + 2\delta_{Bb}\alpha_e - \delta_{bb}\alpha_e$$

$$(dd)_{be} = \delta_{BB}\delta_{EE} - 2\delta_{BB}\delta_{Ee} + \delta_{BB}\delta_{ee} - 2\delta_{Bb}\delta_{EE} + 4\delta_{Bb}\delta_{Ee} - 2\delta_{Bb}\delta_{ee} + \delta_{bb}\delta_{EE} - 2\delta_{bb}\delta_{Ee} + \delta_{bb}\delta_{ee}$$

where, for example, $\alpha_B\alpha_E$ is the additive × additive epistatic effect between the allele B of QTL 1 and the allele E of QTL 2. The eight reduced models are associated with the following null hypotheses: (1) H_0: no QTL in LD with the markers; (2) H_0: no QTL in LD with SNP 2 and no dominance for QTL 1; (3) H_0: no QTL in LD with SNP 2; (4) H_0: no dominance for QTL 2; (5) no epistasis; (6) H_0: only additive × additive epistatic effects; (7) H_0: only additive × additive and additive × dominance epistatic effects; (8) H_0: no dominance × dominance epistatic effect.

The alternative model below has been fitted for genomic prediction including dominance and epistatic effects:

$$G = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \beta_3 x_3 + \beta_4 x_4 + \beta_5 x_1 x_3 + \beta_6 x_1 x_4 + \beta_7 x_2 x_3 + \beta_8 x_2 x_4 + error$$

where x_1, x_3 = 1, 0, or –1 if the individual is homozygous for an SNP allele (C or F), heterozygous, or homozygous for the other SNP allele (c or f), respectively, and x_2, x_4 = 0 or 1 if the individual is homozygous or heterozygous, respectively. Using the same procedure described, it can be demonstrated that the only differences relative to the previous model is that $\beta_2 = d_{SNP1}$, $\beta_4 = d_{SNP2}$, $\beta_6 = -(\alpha\delta)_{SNP1,SNP2}$, and $\beta_7 = -(\delta\alpha)_{SNP1,SNP2}$. Thus, both models are equivalent. Notice that for this second model, the individual SNP genotypic values are defined as $G_{CC} = m_c + a_c$, $G_{Cc} = m_c + d_c$, and $G_{cc} = m_c - a_c$, where the SNP parameters are $m_c = M + (q_c - p_c)\,\alpha_{\mathrm{SNP}} - (1 - 2\mathrm{p_c\ q_c})d_{SNP}$, $a_c = \alpha_{SNP} - (q_c - p_c)d_{SNP}$ and $d_c = d_{SNP}$.

We should also highlight the fact that extension of the Kempthorne's model for SNP effects (Mao et al., 2006) is a third alternative model for genomic prediction. However, compared to previous models, Kempthorne's model is much less parsimonious. Including first degree epistasis for the SNPs (involving pairs of SNPs) increases the number of effects to be predicted by four times (9 versus 36 effects, including the population mean).

SNP effects and variances

The parameters M, α_{SNP} and d_{SNP} are the population mean, the average effect of a SNP substitution, and the dominance deviation at a SNP locus, respectively. The other parameters are the epistatic effects for the SNPs. Notice the relationship between marker effect and marker frequency, LD value, and average effect of QTL substitution, QTL dominance deviation, or QTL epistatic effects. As highlighted by Cockerham (1954), unless all linear combinations of epistatic effects (aa, ad, da, and dd) are zero, there is epistasis between the QTLs. Importantly, these linear combinations of epistatic effects are equivalent to the linear combinations of genotypic values presented by Cockerham (1954) and Viana (2004b).

Thus, a whole-genome analysis provides the population mean, the average effects of SNP substitution, the SNP dominance deviations, and the SNP epistatic effects. The SNP additive, dominance, and epistatic variances are the sums of squares of the linear, quadratic, linear × linear, linear × quadratic, quadratic × linear, and quadratic × quadratic effects, respectively, as demonstrated below:

$$R(\beta_1|\beta_0) = 2p_c q_c \alpha_{SNP1}^2 = \sigma_{A(SNP1)}^2$$

$$R(\beta_2|\beta_0,\beta_1) = 4p_c^2 q_c^2 d_{SNP1}^2 = \sigma_{D(SNP1)}^2$$

$$R(\beta_3|\beta_0,\beta_1,\beta_2) = 2p_f q_f \alpha_{SNP2}^2 = \sigma_{A(SNP2)}^2$$

$$R(\beta_4|\beta_0,\beta_1,\beta_2,\beta_3) = 4p_f^2 q_f^2 d_{SNP2}^2 = \sigma_{D(SNP2)}^2$$

$$R\left(\beta_5|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4\right) = 4p_cq_cp_fq_f\kappa_{bc}^2\kappa_{ef}^2(aa)_{be}^2 = \sigma^2_{AA(SNP1,SNP2)}$$

$$R\left(\beta_6|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5\right) = 2p_cq_cp_f^2q_f^2\kappa_{bc}^2\kappa_{ef}^4(ad)_{be}^2 = \sigma^2_{AD(SNP1,SNP2)}$$

$$R\left(\beta_7|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6\right) = 2p_c^2q_c^2p_fq_f\kappa_{bc}^4\kappa_{ef}^2(da)_{be}^2 = \sigma^2_{DA(SNP1,SNP2)}$$

$$R\left(\beta_8|\beta_0,\beta_1,\beta_2,\beta_3,\beta_4,\beta_5,\beta_6,\beta_7\right) = p_c^2q_c^2p_f^2q_f^2\kappa_{bc}^4\kappa_{ef}^4(dd)_{be}^2 = \sigma^2_{DD(SNP1,SNP2)}$$

where R(.|.) is a difference between two nested R(.) terms with the additional effect stated before the vertical bar and the effect(s) common to both models after the bar, $\sigma^2_{A(SNP)}$ is the SNP additive variance, $\sigma^2_{D(SNP)}$ is the SNP dominance variance, and $\sigma^2_{AA(SNP1,SNP2)}$, $\sigma^2_{AD(SNP1,SNP2)}$, $\sigma^2_{DA(SNP1,SNP2)}$ and $\sigma^2_{DD(SNP1,SNP2)}$ are the epistatic variances for the SNPs.

Extension for two QTLs in LD with a SNP

Assume now one SNP (SNP 1; alleles C/c) in LD with two QTLs (QTL 1 and QTL 2; alleles B/b and E/e), a second SNP (SNP 2; alleles F/f) in LD with a third QTL (QTL 3; alleles H/h), and (for simplicity) that SNP 1 and QTLs 1 and 2 are in linkage equilibrium in relation to SNP 2 and QTL 3. Using the same procedure previously described, it can demonstrated that:

$$\beta_0 = M$$

$$\beta_1 = \left[\frac{\Delta_{bc}^{(-1)}}{p_cq_c}\right]\alpha_b + \left[\frac{\Delta_{ce}^{(-1)}}{p_cq_c}\right]\alpha_e = \kappa_{bc}\alpha_b + \kappa_{ce}\alpha_e = \alpha_{SNP1}$$

$$\beta_2 = -\left(\kappa_{bc}^2d_b + \kappa_{ce}^2d_e\right) = -d_{SNP1}$$

$$\beta_3 = \left[\frac{\Delta_{fh}^{(-1)}}{p_fq_f}\right]\alpha_h = \kappa_{fh}\alpha_h = \alpha_{SNP2}$$

$$\beta_4 = -\kappa_{fh}^2d_h = -d_{SNP2}$$

$$\beta_5 = \kappa_{bc}\kappa_{fh}(aa)_{bh} + \kappa_{ce}\kappa_{fh}(aa)_{eh} = (\alpha\alpha)_{SNP1,SNP2}$$

$$\beta_6 = \left(1/2\right)\kappa_{bc}\kappa_{fh}^2(ad)_{bh} + \left(1/2\right)\kappa_{ce}\kappa_{fh}^2(ad)_{eh} = (\alpha\delta)_{SNP1,SNP2}$$

$$\beta_7 = \left(1/2\right)\kappa_{bc}^2\kappa_{fh}(da)_{bh} + \left(1/2\right)\kappa_{ce}^2\kappa_{fh}(da)_{eh} = (\delta\alpha)_{SNP1,SNP2}$$

$$\beta_8 = \left(1/4\right)\kappa_{bc}^2\kappa_{fh}^2(dd)_{bh} + \left(1/4\right)\kappa_{ce}^2\kappa_{fh}^2(dd)_{eh} = (\delta\delta)_{SNP1,SNP2}$$

Relationship between SNP dominance value and QTL dominance value

Assuming a SNP and a QTL in LD, the dominance values relative to the SNP are proportional to the dominance values relative to the QTL, as shown below for SNP 1 and QTL 1:

$$D_{CC} = \left(q_c/q_b\right)^2\kappa_{bc}^2D_{BB} = -\left(q_c^2/p_bq_b\right)\kappa_{bc}^2D_{Bd} = \left(q_c/p_b\right)^2\kappa_{bc}^2D_{bb}$$

$$D_{Cc} = -\left(p_cq_c/q_b^2\right)\kappa_{bc}^2D_{BB} = \left(p_cq_c/p_bq_b\right)\kappa_{bc}^2D_{Bb} = -\left(p_cq_c/p_b^2\right)\kappa_{bc}^2D_{bb}$$

$$D_{cc} = \left(p_c/q_b\right)^2\kappa_{bc}^2D_{BB} = -\left(p_c^2/p_bq_b\right)\kappa_{bc}^2D_{Bd} = \left(p_c/p_b\right)^2\kappa_{bc}^2D_{bb}$$

where the SNP dominance values are $D_{CC} = -2q_c^2d_{SNP1}$, $D_{Cc} = 2p_cq_cd_{SNP1}$, and $D_{bb} = -2p_c^2d_{SNP1}$, and the QTL dominance values are $D_{BB} = -2q_b^2d_b$, $D_{Bb} = 2p_bq_bd_b$, and $D_{bb} = -2p_b^2d_b$. Notice the relationship between the SNP dominance value and marker and QTL frequency, LD value and dominance genetic value.

Accuracy of dominance value prediction

Based on the previous results, a predictor of the QTL dominance value is the SNP dominance value, i.e., $\tilde{D}_{QTL}^1 = D_{SNP} = u_1d_{SNP}$ (equation 3), where $u_1 = -2q_c^2$ for SNP genotype CC, $u_1 = 2p_cq_c$ for Cc, or $u_1 = -2p_c^2$ for cc. However, the predictor that has been used in most of the whole-genome analysis of field and simulated data for predicting genotypic value is $\tilde{D}_{QTL}^2 = u_2d_{SNP}$, where $u_2 = 0$ for SNP genotype CC, $u_2 = 1$ for Cc, or $u_2 = 0$ for cc. These predictors have the same covariance with the dominance value for the QTL, given by:

$$\begin{aligned}Cov\left(D_{QTL},\tilde{D}_{QTL}^1\right) &= f_{22}\left(-2q_b^2d_b\right)\left(-2q_c^2d_{SNP1}\right) + \ldots + f_{11}\left(2p_bq_bd_b\right)\left(2p_cq_cd_{SNP1}\right)\\ &+ \ldots + f_{00}\left(-2p_b^2d_b\right)\left(-2p_c^2d_{SNP1}\right) - 0.0 = 4p_c^2q_c^2d_{SNP1}^2 = \sigma^2_{D(SNP1)}\\ &= Cov\left(D_{QTL,}\tilde{D}_{QTL}^2\right)\end{aligned}$$

However, the two predictors have different variances given by:

$$\begin{aligned}Var\left(\tilde{D}_{QTL}^1\right) &= f_{.2}\left(-2q_c^2d_{SNP1}\right)^2 + f_{.1}\left(2p_cq_cd_{SNP1}\right)^2 + f_{.0}\left(-2p_c^2d_{SNP1}\right)^2 - 0^2\\ &= \sigma^2_{D(SNP1)}\end{aligned}$$

$$\begin{aligned}Var\left(\tilde{D}_{QTL}^2\right) &= f_{.2}\left(0\right)^2 + f_{.1}\left(d_{SNP1}\right)^2 + f_{.0}\left(0\right)^2 - \left(2p_cq_cd_{SNP1}\right)^2\\ &= \left(\frac{1}{2p_cq_c} - 1\right)\sigma^2_{D(SNP1)}\end{aligned}$$

Thus, the predictor of greater accuracy (lower variance) is $\tilde{D}_{QTL}^1$. The predictors $\tilde{D}_{QTL}^1$ and $\tilde{D}_{QTL}^2$ have the same accuracy when the SNP allelic frequencies are equal (because $\left(1/2p_bq_b\right) - 1 = 1$). The accuracy (correlation between the dominance value for the QTL and the value predicted by the SNP) of predictor 1 is:

$$\rho_{D_{QTL},\tilde{D}_{QTL}^1} = \frac{\sigma^2_{D(SNP1)}}{\sqrt{\sigma^2_{D(QTL1)}\sigma^2_{D(SNP1)}}}$$

where $\sigma^2_{D(QTL1)} = 4p_b^2q_b^2d_b^2$ is the QTL 1 dominance variance.

Dominance value predictor

Generalizing, the dominance genetic value relative to k QTLs predicted by s SNPs is $\tilde{D}^1 = \sum_{r=1}^{s} u_{1(r)}d_{SNP(r)}$. The accuracy of predictor 1 is:

$$\rho_{D_{QTL},\tilde{D}_{QTL}^1} = \sum_{r=1}^{s}\sigma^2_{D(SNP(r))} / \sqrt{\sigma^2_D\sigma^2_{D(SNP)}}$$

where: $\sigma^2_{D(SNP(r))} = 4p_r^2q_r^2d_{SNP(r)}^2$ is the dominance variance for the SNP r,

$$d_{SNP(r)} = \sum_{i=1}^{k'}\left[\frac{\Delta_{ri}^{(-1)}}{p_rq_r}\right]^2 d_i = \sum_{i=1}^{k'}\kappa_{ri}^2d_i$$

is the SNP dominance deviation (k' is the number of QTLs in LD with the SNP r),

$$\sigma_D^2 = 4\sum_{i=1}^{k} p_i^2 q_i^2 d_i^2 + 8\sum_{i=1<}^{k-1}\sum_{j=2}^{k}\left[\Delta_{ij}^{(-1)}\right]^2 d_i d_j$$

(Viana, 2004a) is the dominance variance, and

$$\sigma_{D(SNP)}^2 = 4\sum_{r=1}^{s} p_r^2 q_r^2 d_{SNP(r)}^2 + 8\sum_{r=1<}^{s-1}\sum_{t=2}^{s}\left[\Delta_{rt}^{(-1)}\right]^2 d_{SNP(r)} d_{SNP(t)}$$

(Viana, 2004a) is the variance of the dominance value predictor (dominance genomic value variance).

The best genotypic value predictor is $\tilde{G} = \hat{M} + \tilde{A} + \tilde{D}^1$ (assuming absence of epistasis), where $\hat{M}$ is the estimator of the population mean and $\tilde{A}$ is the predictor of the breeding value.

Prediction of epistatic values

Based on the previous simplified additive-dominance with epistasis model (QTL 1 and SNP 1 in LD, QTL 2 and SNP 2 in LD, and QTL 1 and SNP 1 in linkage equilibrium relative to QTL 2 and SNP 2), we assessed the prediction accuracy of the additive × additive epistatic value for a predictor based on Kempthorne (1954), given by:

$$\tilde{A}A = AA_{rstu} = \alpha_r\alpha_t + \alpha_r\alpha_u + \alpha_s\alpha_t + \alpha_s\alpha_u$$

where $\tilde{A}A$ is the additive × additive genetic value predictor of the individual with SNP genotype rstu, r and s are the alleles for the first SNP, t and u are the alleles for the second SNP, and A_{rstu} is the additive × additive value for the SNPs. The SNP additive × additive effect is:

$$\alpha_r\alpha_t = \bar{G}_{r.t.} - \bar{G}_{r...} - \bar{G}_{..t.} + \bar{G}_{....}$$

where $\bar{G}_{r.t.}$ is the mean of the individuals with alleles r and t, $\bar{G}_{r...}$ the mean of the individuals with allele r, $\bar{G}_{..t.}$ the mean of the individuals with allele t, and $\bar{G}_{....}$ is the population mean.

Assuming distance SNP 1 - QTL 1 = 0.0002 cM, distance SNP 2 - QTL 2 = 0.0002 cM, p_b = 0.4662, p_c = 0.6241, p_e = 0.4914, p_f = 0.4460, a_b = 0.6672, a_e = 0.6435, degree of dominance = 1.0000, and duplicate recessive epistasis (complementary gene action), we have that Δ_{bc} = 0.1154, Δ_{ef} = 0.1273, α_{SNP1} = 0.3502, α_{SNP2} = 0.3374, d_{SNP1} = 0.1613, d_{SNP2} = 0.1709, $(\alpha\alpha)_{\mathrm{SNP1,SNP2}}$ = 0.1238, $(\alpha\delta)_{SNP1,SNP2}$ = –0.0627, $(\delta\alpha)_{\mathrm{SNP1,SNP2}}$ = –0.0570, and $(\delta\delta)_{SNP1,SNP2}$ = 0.0289. The accuracy of the additive × additive genetic value was 0.2232. This accuracy corresponds to the correlation between the SNP additive × additive values (nine values, computed based on Kempthorne (1954)) and the QTL additive × additive values (nine parametric values, computed from Kempthorne (1954)). Thus, we correlated 81 values, weighted by the genotype probabilities.

The predictors of the additive × dominance, dominance × additive, and dominance × dominance epistatic values can also be defined based on Kempthorne (1954). Thus,

$$\tilde{A}D = AD_{rstu} = \alpha_r\delta_{tu} + \alpha_s\delta_{tu}$$

$$\tilde{D}D = DD_{rstu} = \delta_{rs}\delta_{tu}$$

where:

$$\alpha_r\delta_{tu} = \bar{G}_{r.tu} + \bar{G}_{r...} + \bar{G}_{..t.} + \bar{G}_{...u} - \bar{G}_{r.t.} - \bar{G}_{r..u} - \bar{G}_{..tu} - \bar{G}_{....}$$

$$\begin{aligned}\delta_{rs}\delta_{tu} = {} & G_{rstu} - \bar{G}_{....} - (\alpha_r + \alpha_s) - (\alpha_t + \alpha_u) - \delta_{rs} - \delta_{tu} \\ & - (\alpha_r\alpha_t + \alpha_r\alpha_u + \alpha_s\alpha_t + \alpha_s\alpha_u) - (\alpha_r\delta_{tu} + \alpha_s\delta_{tu}) - (\delta_{rs}\alpha_t + \delta_{rs}\alpha_u)\end{aligned}$$

$$\alpha_r = \bar{G}_{r...} - \bar{G}_{....}$$

$$\delta_{rs} = \bar{G}_{rs..} - \bar{G}_{r...} - \bar{G}_{.s..} + \bar{G}_{....}$$

Based on the estimated prediction accuracies of additive, dominance, and epistatic values on this simplified scenario, we can state that genomic prediction of epistatic values should be less accurate than genomic prediction of dominance and additive values. However, because the prediction accuracies of dominance and epistatic values are positive, inclusion of non-additive effects should increase the prediction accuracy of genotypic value in an inbred population (because the genetic values are not independent).

Bias in the prediction of epistatic values

The additive-dominance with epistasis model for genomic selection has a limitation for predicting epistatic values based on markers, owing to LD being between a QTL and two or more markers. That is, the linear × linear, linear × quadratic, quadratic × linear; and quadratic × quadratic effects and variances are not necessarily nil in the absence of epistasis. Assuming a QTL (alleles B/b) in LD with two SNPs (alleles C/c and E/e), distance SNP 1 - QTL = 0.0002 cM, distance QTL - SNP 2 = 0.0003 cM, complete interference, p_c = 0.4073, p_b = 0.4960, p_e = 0.4115, a_b = 4.5 and degree of dominance = –0.9958, we have that Δ_{bc} = –0.0495, Δ_{be} = –0.0801, Δ_{ce} = –0.0786, α_{SNP1} = –0.9153, α_{SNP2} = –1.4764, d_{SNP1} = –0.1884, d_{SNP2} = –0.4902, and $(\alpha\alpha)_{SNP1,SNP2}$ = 0.5155.

Simulation

The data set was simulated using the REALbreeding program (available on request), which is under development by the first author using the REALbasic software. A total of 5000 SNPs and 100 QTLs were distributed on 10 chromosomes with 500 SNPs and 10 QTLs *per* chromosome, covering 50 cM on average (density of one SNP each 0.1 cM on average). The QTLs were distributed in the regions covered by markers. Next, the software simulated a Hardy-Weinberg equilibrium population with LD. This population was a composite, generated by crossing two populations in linkage equilibrium followed by a generation of random crosses. The number of plants was 500 (effective population size of 1000).

Finally, based on user input, the software computed all genetic parameters and LD values between QTLs. The input information includes minimum and maximum genotypic values for homozygotes (allowing for computation of the parameter *a* for each QTL), degree of dominance (d/a), direction of dominance, and the broad sense heritability. The REALbreeding program saves two main files, one with the marker genotypes and the other with the additive, dominance, and phenotypic values (the current version does not compute epistatic values). The true additive and dominance genetic values and variances are computed from the population gene frequencies (random values), LD values, average effects of gene substitution, and dominance deviations. The phenotypic values are computed from the true population mean, additive and dominance values, and from error effects sampled from a normal distribution. The error variance is computed from the broad sense heritability. The LD in a composite is

$$\Delta_{ab}^{(-1)} = \left(\frac{1-2\theta_{ab}}{4}\right)\left(p_a^1 - p_a^2\right)\left(p_b^1 - p_b^2\right),$$

where θ_{ab} is the frequency of recombinant gametes and the indices 1 and 2 refer to the gene frequencies (p) in the parental populations.

We considered two popcorn traits, three SNP densities, two heritabilities, two sample sizes, and two populations showing LD, totaling 48 scenarios. For each scenario, 50 simulations were carried out. The minimum and maximum genotypic values of homozygotes for grain yield and expansion volume were 20 and 200 g per plant, and 5 and 50 mL g^{-1}. Positive unidirectional dominance ($0 < (d/a)_i \leq 1.2$) was assumed for grain yield and bidirectional dominance ($-1.2 \leq (d/a)_i \leq 1.2$) was assumed for expansion volume (i = 1, 2, ..., 100). The other SNP densities, one marker every cM and one marker every10 cM on average, were obtained by random choices of 51 and 6 SNPs by chromosome, respectively, also using the REALbreeding software. The broad sense heritabilities were 0.3 and 0.7. Thus, the accuracies of the phenotypic values were 0.548 and 0.837. The sample size was 500 or 200 individuals genotyped and phenotyped. The second population with LD was obtained from the composite (generation 0) after five generations of random crosses (generation 5).

Regarding the SNPs, the averages of the absolute LD values (the difference between gametic frequencies observed and expected under linkage equilibrium) in the LD blocks in generation 0, were 0.0978, 0.1073, and 0.1413 for the densities 10, 1, and 0.1 cM, respectively. The values for generation 5 were 0.0581, 0.0565, and 0.0901, respectively. The LD blocks were defined based on an LOD (logarithm [base 10] of odds) score of 3 (to declare LD for two linked SNPs), also using the REALbreeding software. The corresponding r^2 (the square of the correlation coefficient between the alleles of two loci) values were 0.1785, 0.2089, and 0.4079 for generation 0, and 0.0843, 0.0829, and 0.2420 for generation 5. All data used in this paper are accessible on request.

Statistical analysis of the simulated data

The method used for genomic selection was RR-BLUP (ridge regression best linear unbiased prediction). For the analyses we used the *rrBLUP* (Endelman, 2011). The accuracies of dominance and genotypic value prediction were obtained by the correlation between the true values computed by REALbreeding and the values predicted by RR-BLUP.

Results

Theoretical results

We show that the predictor of dominance value is proportional to the square of the LD value and to the dominance deviation for each QTL that is in LD with each marker (see $\tilde{D}^1$ and $d_{SNP(r)}$). The dominance value predictor of greater accuracy (lower variance) is that weighted by the SNP frequencies (compare the variances of $\tilde{D}^1$ and $\tilde{D}^2$). The linear × linear, linear × quadratic, quadratic × linear, and quadratic × quadratic SNP effects are proportional to the corresponding linear combinations of epistatic effects for QTLs and the LD values between SNPs and QTLs (see regression coefficients β_5 to β_8). Linkage disequilibrium between two SNPs with a common QTL results in a bias in the prediction of epistatic values. The SNP epistatic values are proportional to the corresponding epistatic genetic values.

Simulation results

Regardless of generation, sample size, SNP density, and heritability, the prediction accuracy of the dominance genetic value was, with one exception, higher for grain yield, indicating that the prediction is less accurate when dominance is bidirectional (Table 1). The ratio between the accuracies of grain yield and expansion volume ranged from 1.0 to 3.1, with the ratio being inversely proportional to heritability. Regardless of the other factors, the accuracy of dominance value prediction was proportional to the SNP density. On average, the increase from 1 SNP every 10 cM to 1 SNP every cM and from 1 SNP every cM to 1 SNP every 0.1 cM determined increments in the accuracy of 114 and 46 %, respectively. The increments were higher for grain yield under high heritability. The decrease in sample size from 500 to 200 caused a decrease in dominance value prediction accuracy, regardless of the other factors, although the decreases were of reduced magnitude. The average decreases ranged from 1 to 12 % and were also inversely proportional to the heritability. The decrease in LD with five generations of random mating also caused a decrease, generally of low magnitude, in dominance value prediction accuracy that is inversely proportional to the SNP density and heritability. The average decreases ranged from 0 to 51 %. For the two traits and regardless of the other factors, increasing heritability led to an increase in dominance value prediction accuracy, ranging from 16 to 100 %.

Table 1 – Prediction accuracy of dominance value and its standard deviation, for expansion volume and grain yield, regarding two accuracy levels of the phenotypic value, three SNP (single nucleotide polymorphisms) densities, two sample sizes, and two generations.

Gen.	Sample	SNP density (cM)	Accuracy of the phenotypic value			
			0.548		0.837	
			Expansion volume	Grain yield	Expansion volume	Grain yield
0	200	10	0.082 ± 0.10	0.204 ± 0.10	0.146 ± 0.08	0.280 ± 0.09
		1	0.144 ± 0.11	0.423 ± 0.10	0.288 ± 0.07	0.602 ± 0.07
		0.1	0.210 ± 0.11	0.557 ± 0.08	0.401 ± 0.07	0.726 ± 0.04
	500	10	0.095 ± 0.07	0.238 ± 0.07	0.139 ± 0.06	0.335 ± 0.06
		1	0.168 ± 0.08	0.525 ± 0.07	0.301 ± 0.06	0.663 ± 0.03
		0.1	0.256 ± 0.09	0.661 ± 0.05	0.437 ± 0.05	0.764 ± 0.02
5	200	10	0.085 ± 0.07	0.093 ± 0.11	0.125 ± 0.09	0.144 ± 0.10
		1	0.165 ± 0.08	0.200 ± 0.09	0.288 ± 0.08	0.358 ± 0.08
		0.1	0.220 ± 0.09	0.341 ± 0.09	0.400 ± 0.08	0.539 ± 0.07
	500	10	0.089 ± 0.05	0.088 ± 0.07	0.125 ± 0.05	0.136 ± 0.06
		1	0.150 ± 0.06	0.226 ± 0.07	0.265 ± 0.05	0.387 ± 0.05
		0.1	0.242 ± 0.07	0.417 ± 0.07	0.440 ± 0.05	0.605 ± 0.04

Table 2 – Prediction accuracy of genotypic value and its standard deviation, for expansion volume and grain yield, regarding two accuracy levels of the phenotypic value, three SNP (single nucleotide polymorphisms) densities, two sample sizes, and two generations.

Gen.	Sample	SNP density (cM)	Accuracy of the phenotypic value			
			0.548		0.837	
			Expansion volume	Grain yield	Expansion volume	Grain yield
0	200	10	0.443 ± 0.05	0.469 ± 0.06	0.574 ± 0.04	0.574 ± 0.04
		1	0.598 ± 0.06	0.643 ± 0.06	0.778 ± 0.03	0.793 ± 0.02
		0.1	0.639 ± 0.06	0.681 ± 0.05	0.826 ± 0.02	0.815 ± 0.03
	500	10	0.453 ± 0.04	0.452 ± 0.07	0.532 ± 0.03	0.546 ± 0.03
		1	0.618 ± 0.03	0.661 ± 0.04	0.735 ± 0.02	0.745 ± 0.02
		0.1	0.663 ± 0.03	0.697 ± 0.03	0.804 ± 0.02	0.771 ± 0.02
5	200	10	0.343 ± 0.07	0.340 ± 0.07	0.484 ± 0.05	0.491 ± 0.06
		1	0.536 ± 0.07	0.536 ± 0.07	0.759 ± 0.04	0.759 ± 0.04
		0.1	0.580 ± 0.07	0.595 ± 0.07	0.820 ± 0.03	0.822 ± 0.03
	500	10	0.337 ± 0.05	0.330 ± 0.09	0.426 ± 0.04	0.447 ± 0.04
		1	0.551 ± 0.04	0.564 ± 0.04	0.707 ± 0.02	0.722 ± 0.02
		0.1	0.621 ± 0.04	0.653 ± 0.04	0.816 ± 0.02	0.808 ± 0.02

The degree of dominance did not affect genotypic value prediction accuracy and the approach to maximum accuracy is asymptotic with the increase in SNP density (Table 2). The ratio between the accuracies of the traits ranged from 1.0 to 1.1 and, in general, an increase of relevant magnitude occurred only in genotypic value prediction accuracy when increasing the density from 1 SNP every 10 cM to 1 SNP every cM (49 % on average). The average increase in accuracy when increasing the density from 1 SNP every cM to 1 SNP every 0.1 cM was 9 % (maximum of 16 %). Reducing the sample size from 500 to 200 did not affect the magnitude of the accuracy of genotypic value prediction, with a maximum decrease in magnitude of 4 %. Also the decrease in LD after 5 generations of random mating also did not considerably decrease genotypic value prediction accuracy, regardless of the other factors. The decreases ranged from 0 to 26 %. Also, regardless of generation, sample size, SNP density, and trait, the accuracy of genotypic value prediction was proportional to heritability. The increases ranged from 11 to 44 %.

Assuming absence of epistasis, the accuracy of breeding value prediction by fitting the additive, additive-dominance and additive-dominance with additive × additive epistasis models were equivalent, regardless of heritability and SNP density (Table 3). The same was found for the accuracy of dominance value prediction by fitting the additive-dominance and additive-dominance with additive × additive epistasis models.

Discussion

The prediction of breeding values based on markers is relevant to both animal and plant population improvement. The selection of couples in animal breeding (Toro and Varona, 2010), of clones in forestry breeding (Denis and Bouvet, 2013), and of hybrids in annual crop breeding (Zhao et al., 2013) are examples

Table 3 – Accuracy of additive (A), dominance (D), and genotypic (G) values by fitting the additive-dominance (AD) model and the additive-dominance with additive × additive epistasis (ADE) model, assuming no epistasis, generation 0, two accuracy levels of the phenotypic value, two SNP (single nucleotide polymorphisms) densities, and 500 individuals.

Model	Value	SNP density (cM)	Accuracy of the phenotypic value			
			0.548		0.837	
			Expansion volume	Grain yield	Expansion volume	Grain yield
AD	A	10	0.598 ± 0.03	0.586 ± 0.04	0.649 ± 0.02	0.626 ± 0.03
		1	0.693 ± 0.03	0.685 ± 0.04	0.769 ± 0.02	0.751 ± 0.02
ADE	A	10	0.599 ± 0.03	0.581 ± 0.05	0.640 ± 0.02	0.627 ± 0.03
		1	0.696 ± 0.03	0.693 ± 0.04	0.773 ± 0.02	0.776 ± 0.02
	D	10	0.089 ± 0.06	0.207 ± 0.07	0.114 ± 0.05	0.266 ± 0.06
		1	0.175 ± 0.08	0.548 ± 0.07	0.320 ± 0.06	0.686 ± 0.03
	G	10	0.446 ± 0.04	0.421 ± 0.04	0.522 ± 0.03	0.503 ± 0.03
		1	0.676 ± 0.03	0.666 ± 0.04	0.799 ± 0.02	0.809 ± 0.02

of the relevance of predicting genotypic value based on markers in animal and plant breeding. Zeng et al. (2013) concluded that genomic selection based on the dominance model maximized the cumulative response to selection of purebred animals for crossbred performance. Comparing genomic selection and pedigree-based BLUP by assuming dominance gene action, Denis and Bouvet (2013) observed that genomic selection provided higher gain per unit time for a clone population, mainly in the first selection cycle. Massman et al. (2013) also compared genomic selection and pedigree-based BLUP for prediction of maize single-cross performance for grain yield and other traits. They observed high accuracy in the genome-wide predictions, but the accuracy was inferior to those obtained with pedigree-based BLUP.

Although our simulation study revealed lower accuracy of dominance value prediction with bidirectional dominance, relative to unidirectional dominance, genotypic value prediction follows the same rules governing the breeding value prediction. Under high SNP density (at least 1 SNP every cM), the accuracy of additive and genotypic value predictions are equivalent, regardless of the degree of dominance. As already established for genomic selection based on breeding value prediction, in relation to phenotypic selection, the efficiency of genomic selection based on genotypic value prediction is inversely proportional to heritability. Assuming high SNP density and low heritability, genomic selection is at least as efficient as phenotypic selection. Maximum efficiency can reach 27 %. With regard to traits with high heritability, the efficiency of genomic selection is also high, reaching at least 90 % of phenotypic selection efficiency.

Our results showed as expected because the population analyzed was not inbred (i.e., additive, dominant and epistatic genetic values are independent), that the accuracy of breeding value prediction is equivalent when fitting the additive, additive-dominance, and additive-dominance with epistasis models. It is noteworthy that in our study, in the absence of epistasis and under high SNP density, fitting additive × additive epistatic effects in relation to SNPs led to an increase in the accuracy of genotypic value prediction. Concerning the fitting of the additive-dominance with the epistasis model, several results evidenced an increase in breeding value prediction accuracy with the inclusion of the dominance (Wellmann and Bennewitz, 2012; Technow et al., 2012) and epistatic (Dudley and Johnson, 2009; Long et al., 2010; Wittenburg et al., 2011; Hu et al., 2011; Wang et al., 2012; Su et al., 2012) effects. Carlborg and Haley (2004) emphasized the importance of fitting epistatic effects in complex trait studies. In some genome-wide analyses the proportion of the phenotypic variance explained by epistatic effects ranged from 6 to 83 % (Xu, 2007; Wang et al., 2010, 2012). Interestingly, Xu and Jia (2007) found that whether two markers interact does not depend on whether the loci have individual main effects.

A criticism of the RR-BLUP method is one variance for the additive, dominance and epistatic effects for the SNPs. However the accuracy of additive and genotypic value prediction with the use of the RR-BLUP method based on the likelihood approach is in fully agreement with the accuracy obtained with penalized regression and Bayesian methods. Thus, our results showed that RR-BLUP is a suitable method for genomic selection because it provides at least the same accuracy level of phenotypic selection for traits of low heritability, or at least a slightly lower accuracy level than that of phenotypic selection for traits of high heritability. Using simulated and empirical datasets, Sun et al. (2012) compared the non-parametric methods RKHS (reproducing kernel Hilbert spaces) and pRKHS (which combine supervised principal component analysis and RKHS regression), with RR-BLUP, BayesA and BayesB. Assuming no, low and high epistasis, the non-parametric methods were superior to the other methods in terms of accuracy of additive and genotypic values prediction. Zhao et al. (2013) used RR-BLUP, BayesA, BayesB, BayesC, and BayesCπ to analyze wheat and simulated data assuming the additive with dominance model. The cross validation approach showed slight superiority of RR-BLUP and BayesB regarding the accuracy of predicting the hybrid performance. Interestingly, ignoring dominance effects resulted in equal or even higher prediction accuracies.

As final comments, it is important to highlight the contribution of the theory presented and the relevance of the assessment of genomic prediction of dominance and genotypic values in breeding populations of open-pollinated crops. Since the advent of genomic selection (Meuwissen et al., 2001) some theoretical aspects have been presented (Goddard, 2009; Gianola et al., 2009; Vitezica et al., 2013). However, the theory presented in this paper gives proofs of genomic selection efficacy that have not been fully presented in the previously mentioned relevant papers. Furthermore, based on a simplified scenario including digenic epistasis and in simulated data, we can state that although prediction accuracy of additive value has greater magnitude than the prediction accuracy of dominance and epistatic values, breeders should expect that genome-wide prediction of genotypic value would be as successful as genomic prediction of breeding value, conditional on the choice of adequate SNP density, sample size, and genomic model. Finally, we demonstrated that genomic selection can be successfully applied in recurrent breeding programs for open-pollinated crops, without training population and validation process.

Acknowledgments

We thank the Brazilian National Council for Scientific and Technological Development (CNPq), the Coordination for the Improvement of Higher Level Personnel (Capes), and the Minas Gerais State Foundation for Research Support (Fapemig) for financial support. The second author was supported by the German Federal Ministry of Education and Research (Bonn, Germany) within the AgroClusterEr "Synbreed-Synergistic plant and animal breeding" (Grant 0315526).

References

Carlborg, Ö.; Haley, C.S. 2004. Epistasis: too often neglected in complex trait studies? Nature Reviews 5: 618-625.

Cockerham, C.C. 1954. An extension of the concept of partitioning hereditary variance for analysis of covariances among relatives when epistasis is present. Genetics 39: 859-882.

Daetwyler, H.D.; Calus, M.P.L.; Pong-Wong, R.; De Los Campos, G.; Hickey, J.M. 2013. Genomic prediction in animals and plants: simulation of data, validation, reporting, and benchmarking. Genetics 193: 347-365.

De Los Campos, G.; Hickey, J.M.; Pong-Wong, R.; Daetwyler, H.D.; Calus, M.P.L. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. Genetics 193: 327-345.

Denis, M.; Bouvet, J.M. 2013. Efficiency of genomic selection with models including dominance effect in the context of *Eucalyptus* breeding. Tree Genetics & Genomes 9: 37-51.

Dudley, J.W.; Johnson, G.R. 2009. Epistatic models improve prediction of performance in corn. Crop Science 49: 763-770.

Endelman, J.B. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. The Plant Genome 4: 250-255.

Gianola, D.; De Los Campos, G.; Hill, W.G.; Manfredi, E.; Fernando, R. 2009. Additive genetic variability and the Bayesian alphabet. Genetics 183: 347-363.

Goddard, M. 2009. Genomic selection: prediction of accuracy and maximization of long term response. Genetica 136: 245-257.

Hill, W.G.; Robertson, A. 1968. Linkage disequilibrium in finite populations. Theoretical and Applied Genetics 38: 226-231.

Hu, Z.; Li, Y.; Song, X.; Cai, X.; Xu, S.; Li, W. 2011. Genomic value prediction for quantitative traits under the epistatic model. BMC Genetics 12: 15.

Kempthorne, O. 1957. An Introduction to Genetic Statistics. Iowa State University Press, Ames, IA, USA.

Kempthorne, O. 1954. The theoretical values of correlations between relatives in random mating populations. Genetics 40: 153-167.

Long, N.; Gianola, D.; Rosa, G.J.M.; Weigel, K.A.; Kranis, A.; González-Recio, O. 2010. Radial basis function regression methods for predicting quantitative traits using SNP markers. Genetics Research 92: 209-225.

Mao, Y.; London, N.R.; Ma, L.; Dvorkin, D.; Da, Y. 2006. Detection of SNP epistasis effects of quantitative traits using an extended Kempthorne model. Physiological Genomics 28: 46-52.

Massman, J.M.; Gordillo, A.; Lorenzana, R.E.; Bernardo, R. 2013. Genomewide predictions from maize single-cross data. Theoretical and Applied Genetics 126: 13-22.

Su, G.; Christensen, O.F.; Ostersen, T.; Henryon, M.; Lund, M.S. 2012. Estimating additive and non-additive genetic variances and predicting genetic merits using genome-wide dense single nucleotide polymorphism markers. PLoS One 7: e45293.

Sun, X.; Ma, P.; Mumm, R.H. 2012. Nonparametric method for genomics-based prediction of performance of quantitative traits involving epistasis in plant breeding. PLoS One 7: e50604.

Technow, F.; Riedelsheimer, C.; Schrag, T.A.; Melchinger, A.E. 2012. Genomic prediction of hybrid performance in maize with models incorporating dominance and population specific marker effects. Theoretical and Applied Genetics 125: 1181-1194.

Toro, M.A.; Varona, L. 2010. A note on mate allocation for dominance handling in genomic selection. Genetics Selection Evolution 42: 33.

Viana, J.M.S. 2004a. Quantitative genetics theory for non-inbred populations in linkage disequilibrium. Genetics and Molecular Biology 27: 594-601.

Viana, J.M.S. 2004b. Relative importance of the epistatic components of genotypic variance in non-inbred populations. Crop Breeding and Applied Biotechnology 4: 18-27.

Vitezica, Z.G.; Varona, L.; Legarra, A. 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. Genetics 195: 1223-1230.

Xu, S. 2007. An empirical Bayes method for estimating epistatic effects of quantitative trait loci. Biometrics 63: 513-521.

Xu, S.; Jia, Z. 2007. Genomewide analysis of epistatic effects for quantitative traits in barley. Genetics 175: 1955-1963.

Wang, D.; El-Basyoni, I.S.; Baenziger, P.S.; Crossa, J.; Eskridge, K.M.; Dweikat, I. 2012. Prediction of genetic values of quantitative traits with epistatic effects in plant breeding populations. Heredity 109: 313-319.

Wang, D.; Eskridge, K.M.; Crossa, J. 2010. Identifying QTLs and epistasis in structured plant populations using adaptive mixed LASSO. Journal of Agricultural, Biological, and Environmental Statistics 16: 170-184.

Wellmann, R.; Bennewitz, J. 2012. Bayesian models with dominance effects for genomic evaluation of quantitative traits. Genetics Research 94: 21-37.

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 Genetics 12: 74.

Zeng, J.; Toosi, A.; Fernando, R.L.; Dekkers, J.C.M.; Garrick, D.J. 2013. Genomic selection of purebred animals for crossbred performance in the presence of dominant gene action. Genetics Selection Evolution 45: 11.

Zhao, Y.; Zeng, J.; Fernando, R.; Reif, J.C. 2013. Genomic prediction of hybrid wheat performance. Crop Science 53: 802-810.