Euphytica (2012) 187:263–276
DOI 10.1007/s10681-012-0722-0

Genome-wide selection in cassava

Eder Jorge de Oliveira · Marcos Deon Vilela de Resende · Vanderlei da Silva Santos · Cláudia Fortes Ferreira · Gilmara Alvarenga Fachardo Oliveira · Maiane Suzarte da Silva · Luciana Alves de Oliveira · Carlos Ivan Aguilar-Vildoso

Received: 26 March 2012 / Accepted: 17 May 2012 / Published online: 4 June 2012

Abstract The main objective of this study was to estimate the selection accuracy and to predict the genetic gain in cassava breeding using genomic selection methodologies. We evaluated 358 cassava genotypes for the following traits: shoot weight (SW), fresh root yield (FRY), starch fraction amylose content (AC), dry matter content (DMC), and starch yield (S-Y). Genotyping was performed using 390 single nucleotide polymorphisms (SNPs), which were used as covariates in the random regression-best linear unbiased prediction model for genomic selection. The heritability values detected by markers for the SW, FRY, AC, DMC, and S-Y traits were 0.25, 0.25, 0.03, 0.20, and 0.26, respectively. Because the low heritability detected for AC, this trait was eliminated from further analysis. Using only the most informative SNPs (118, 92, 56, and 97 SNPs for SW, FRY, DMC, and S-Y, respectively) we observed higher selection accuracy which were 0.83, 0.76, 0.67, and 0.77, respectively to SW, FRY, DMC, and S-Y. With these levels of accuracy and considering a selection cycle reduced by half the time, the theoretical gains with genomic selection compared to phenotypic selection for DMC, FRY, and SW would be 39.42 %, 56.90 %, and 73.96 %, respectively. These results indicate that in the cassava, genomic selection can substantially speed up selection cycles, thereby increasing gains per unit time. Although there are high expectations for incorporating this strategy into breeding programs, we still need to validate the model for other traits and

E. J. de Oliveira (✉) · V. da Silva Santos · C. F. Ferreira · L. A. de Oliveira
Embrapa Cassava and Fruits, Cruz das Almas, BA 44380-000, Brazil
e-mail: eder@cnpmf.embrapa.br

V. da Silva Santos
e-mail: vssantos@cnpmf.embrapa.br

C. F. Ferreira
e-mail: claudiaf@cnpmf.embrapa.br

L. A. de Oliveira
e-mail: luciana@cnpmf.embrapa.br

M. D. V. de Resende
Embrapa Forestry Research, Colombo, PR 83411-000, Brazil
e-mail: marcos.deon@gmail.com

G. A. F. Oliveira · M. S. da Silva
Federal University of Recôncavo da Bahia, Cruz das Almas, BA 44380-000, Brazil
e-mail: gfachardo@yahoo.com.br

M. S. da Silva
e-mail: maisuzarte@yahoo.com.br

C. I. Aguilar-Vildoso
CNPq/Embrapa Cassava and Fruits, Cruz das Almas, BA 44380-000, Brazil
e-mail: vildoso@hotmail.com

evaluate whether the selection accuracy can be improved using more SNPs.

Keywords *Manihot esculenta* Crantz · Breeding · SNP · RR-BLUP

Introduction

Cassava (*Manihot esculenta* Crantz) is a crop of great importance because it is the fourth most important staple food after wheat, rice, and corn, and it is a key component in the diets of millions of people (Nassar and Ortiz 2007). Brazil has approximately two million cassava-cultivated hectares that generate annual revenue of US$ 3.1 billion. Northeastern Brazil produces approximately 36.8 % of the Brazilian national production cultivated within 911,000 ha. Thus, the cassava crop has fundamental importance for the country because it is one of the most relevant commodities for subsistence agriculture and food security. The cassava crop's recent history demonstrates a growing demand for agricultural development as a raw material for multiple industrial applications, including starch yield (S-Y) and starch-derived products, such as alcohol and glucose–fructose syrups (Kunkeaw et al. 2011).

Due to its great social and economic importance, it is necessary to ensure sustainability and competitiveness in cassava production throughout various regions of Brazil. In this sense, plant improvement is one of the areas with the highest rate of return from investments in agricultural research. Cassava genetic improvement has had an impact on the production system, but these impacts are still short of the gains observed for other crops. Indeed, the production of cassava in Brazil from 1961 to 2009 decreased by 0.35 % per year, while cassava yield had gains of only 0.12 % per year (FAO 2011).

The following factors are notable among the many characteristics that contribute to obtaining insignificant yield gains: low agricultural input use, harsh environmental conditions, especially in marginal cultivated zones, and small gains in the yield potential of new varieties. In the latter case, this is partly due to the limited selection strategies used thus far because, typically, most breeding programs promote crosses between contrasting parents from which individuals of the F_1 generation are highly heterozygous. When an improved genotype is identified, it is vegetatively reproduced for validation tests under field conditions. Another important reason for poor progress in genetic improvements of cassava is the lack of public and private investment in breeding.

Because the following conventional strategy is still extensively used in Brazil, the time required for cultivar development is at least 10 years when considering a year for each of the following steps: (a) obtaining F_1 seeds; (b) seedling plant assessment; (c) clonal tests; (d) preliminary tests; and (e) advanced tests. Then, there are at least 2 years for regional assessment trials to estimate adaptability and stability parameters to support material recommendation. In addition, there is another 3 years for reproducing promising clones.

Although conventional breeding still provides genetic gains that justify its use (Ceballos et al. 2004), advanced biological tools have led to increased expectations and maximized gains for several traits of interest, especially those governed by many genes with a small effect. In theory, genotypic information from molecular markers, when associated with phenotypic traits of interest, may be extensively used in order to select individuals with higher genetic value through marker-assisted selection (MAS). Thus, to ensure more significant increases in cassava yield in Brazil and to improve the crop's competitiveness, it is necessary to adopt new biometrics and biotechnology approaches for developing varieties, as well as to place more emphasis on searching for aggregate value in cassava starch, whose traits appear to be controlled by many genes (CIAT 2008).

The prospect of increased selection gains and reduced improvement cycles via MAS culminated with several studies involving QTL (quantitative trait loci) detection in aspects of cassava production. These aspects included root and shoot yield, harvest index, number and diameter of roots, and tuberization rate (Okogbenin and Fregene 2002, 2003); post-harvest physiological deterioration (Cortez et al. 2002); resistance to bacteriosis (Jorge 2000; Jorge et al. 2000, 2001; Wydra et al. 2004; Lopez et al. 2007), and cyanogenic compound content (Kizito et al. 2007; Whankaew et al. 2011). However, the practical application of these QTLs in breeding programs is negligible. According to Dekkers (2004), the main causes of QTLs studies failure is that only a small number of QTLs with major effects are detected, in

contrast with the polygenic nature of the total genetic variation observed for the majority of the quantitative traits. Furthermore, it is necessary to perform large-scale genotyping so that MAS can provide better results than traditional selection.

Advances in genotyping technology enabling large-scale genotyping marker automation, especially for new types of molecular markers, such as SNPs (single nucleotide polymorphisms) (Jenkins and Gibson 2002; Bernardo and Yu 2007), promise to reduce prices per data point for extensive use in various crops. For cassava, SNPs are found at a frequency of one SNP for every 62 base pairs in expressed sequence tags (ESTs) (Lopez et al. 2005). However, despite developing a considerable set of markers, their use is still based only on the crop's defined origin (Olsen 2004), genetic diversity studies (Lopez et al. 2005; Kawuki et al. 2009), and allelic variations in genes related to the carotenoid biosynthetic pathway (Welsch et al. 2010).

The use of large-scale genotyping information in genetic selection strategies is known as genome-wide selection (GWS). According to Meuwissen et al. (2001), under a GWS approach, if a marker is in linkage disequilibrium with the QTL, some alleles of the markers correlate with the positive effects of the QTLs in all families and therefore can be used without the need to establish the linkage phase in each family. Additionally, the combined genotypic and phenotypic data can be used to estimate the genetic merit or predict phenotypic values of the trait of interest.

Genomic selection has been extensively researched and developed for application in animal breeding (Dekkers 2007; Long et al. 2007; Solberg et al. 2008). In plants, this technique also has potential for major impact. Indeed, recent reviews and research with simulated data have demonstrated excellent results using this technology for various crops (Bernardo and Yu 2007; Resende et al. 2008; Wong and Bernardo 2008; Heffner et al. 2009; Zhong et al. 2009; Daetwyler et al. 2010; Meuwissen and Goddard 2010; Villanueva et al. 2011).

In cassava, the prospects for using genomic selection are enormous because the early choice of plants with high genetic/genomic values in segregating populations would already allow selection at the seedling stage. This would save time during the selection process by representing an efficient substitute for phenotypic selection at certain phases of the program, especially for difficult to measure traits or those that require higher precision in data collection. Thus, it is possible to increase the genetic gain per unit time because it is considered that the actual phenotypic value of a cassava clone can only be evaluated in advanced agronomic trials, which occurs at least 5 years after the onset of crossing. In addition, genomic selection has the advantage of applying the results to all families under evaluation; it also has a high accuracy for selection based solely on markers, does not require prior knowledge of QTL positions in linkage maps, avoids biased estimates of the effects of genes or individual QTLs, captures all of the variation due to loci with small effects, and contemplates traits of low heritability with efficiency (Resende et al. 2008).

This study aims to evaluate the use of genomic selection in cassava from a set of cassava accessions (no kinship information) genotyped with SNP molecular markers and evaluate also the efficacy of this procedure in relation to phenotypic selection.

Materials and methods

We evaluated 358 cassava accessions belonging to the germplasm collection of the Embrapa Cassava and Fruit to estimate the genetic effects and validate SNP markers. The genotypes used were hybrids, cultivars and landraces adapted from different Brazilian production regions. This trial was evaluated at Embrapa-Cruz das Almas, Brazil located at 12°48′38″S and 39°6′26″W, and 220 m above sea level in 2010 and 2011. Entries were planted as single rows 9 m long with spacing of 1.0 × 0.9 m between plots and plants using stem cuttings 18 cm long. For the evaluation, we used a randomized block design with three replications and a 10-plant plot. Planting was done in June, and harvesting was done when plants were 12 months old. Irrigation was applied only in the first 2 months and fertilizers were applied as required (Gomes and Silva 2006). None pesticides were used.

Agronomic data

The traits evaluated were as follows: (1) shoot weight (SW), including stems (t ha^{-1}); (2) fresh root yield (FRY) (t ha^{-1}); (3) percent starch fraction amylose content (AC); (4) dry matter content (DMC); and (5)

S-Y by area considering root DMC (S-Y) (t ha^{-1}). All phenotypic values were corrected for the effects of blocks and faults in the experiment.

Genotyping with SNP markers

SNPs were genotyped using the Sequenom iPlex MassARRAY platform, according to the manufacturer's instructions (i.e., based on allele-specific primer extension and resolution by mass spectrometry) (Sequenom, San Diego, California, USA, http://www.sequenom.com/). We genotyped 354 SNP markers derived from gene regions obtained from ESTs for drought tolerance and water productivity, and another 48 derived from the cassava physical map for a total of 402 markers (http://cassava.igs.umaryland.edu/cgi-bin/index.cgi). However, 12 markers were monomorphic and, thus, were not used for genomic selection.

Detecting SNP × phenotype associations

Given the lack of prior information about the association or lack of association between our SNPs with the QTLs involved in expressing the analyzed traits, we initially used all 390 markers for genomic selection. Next, we used genome-wide association studies (GWAS) to assess the association between SNPs and the phenotypic traits via hypothesis testing. Based on Resende (2008), the following regression model based on simple marker was used, which considered the association between the marker and the possible QTL: $y = 1\mu + Xm_i + e$, where y is the vector of phenotypic observations, 1 is the vector with values 1, μ is the scalar reference to the overall average, m_i is the fixed effect of the marker, e refers to the vector of random residuals, and X is the matrix of incidence for m_i, which associates the number of each allele of the SNP to the phenotypes sufficient to fit the effect of only one of the alleles. The GWAS based on single marker regression was done aiming at comparing it with the multiple regression approach done with the GWS approach based on all SNPs as described below.

After implementing the association analysis, we selected the markers that are significantly associated with the phenotypes to be used in the selection, taking into account the comparison of accuracy and reliability estimates in predicting a complete model with all SNPs. All analyses were conducted using the Selegen Genomic RR-BLUP software (Resende 2007), through the REML/BLUP procedure.

Implementing genomic selection

We used the random regression-best linear unbiased prediction (RR-BLUP) method for GWS (Meuwissen et al. 2001), which uses BLUP predictors considering the markers genotypes as random effect covariates (i.e., the phenotypes are regressed based on these covariates). Based on Resende (2008), we used the following linear mixed model to estimate the marker effects:

$y = Xb + Zm + e$, where y is the vector of phenotypic observations, b is the vector of fixed effects, m is the vector of random effect of markers, and e refers to the vector of random residuals. X and Z are incidence matrices for b and m, respectively. Under this model, the following assumptions must be adopted: $m \sim N(0, I\sigma_m^2)$, $E(y)=Xb$, $e \sim N(0, R = I\sigma_e^2)$ and $\mathrm{Var}(y) = V = ZI\sigma_m^2 Z' + R$..

The mixed model equations for predicting m via the RR-BLUP method are equivalent to the following:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + I\frac{\sigma_e^2}{\sigma_g^2/n} \end{bmatrix} \begin{bmatrix} \hat{b} \\ \hat{m} \end{bmatrix} = \begin{bmatrix} X'y \\ Z'y \end{bmatrix},$$

where σ_g^2 refers to the total genetic variance of the trait and σ_e^2 is the residual variance. The overall genetic value of individual j is given by $VGG = \hat{y}_j = \sum_i Z_{ij}\hat{m}_i$, where Z_{ij} equals 0, 1, or 2 according to the allelic dose of the SNPs in the individuals and $i = 1, 2,\ldots,n$ refer-se to the marker i.The prediction equations presented above assumed a priori that all loci explain equal amounts of the genetic variation. Thus, the genetic variation explained by each locus is given by: $\sigma_m^2 = \sigma_g^2/n$, where σ_g^2 is the total genetic variation and n is the number of markers.

Regarding the heritability, Resende et al. (2010) recommend using regression coefficients to assess the adequacy of h^2 to be used to estimate the accuracy and predictive ability of the genomic values, considering that a β value equal to or close to 1 indicates that the prediction between the observed and predicted values was not biased. In contrast, an estimated β value of less than 1 indicates that the genetic values are overestimated and exhibit more variability than expected. Moreover, if the β value is greater than 1, this indicates

that the estimated genetic values exhibit variability below what was expected.

The recovered of the corresponding trait heritability captured by the markers were estimated based on the proportion of the genetic variation explained by the markers using an approach described previously (Gianola et al. 2009; Resende et al. 2010).

Cross-validation was performed by resampling from a group of individuals using the Jackknife procedure, which is based on dividing the N set from the sample data into g groups with sizes equal to k, where $N = gk$. Thus, the cassava accessions were divided into 17 groups of 21 individuals. In each of the 17 analyses, a group was removed from the population and used as a validation population, and the other 336 individuals (belonging to the other 16 groups) were used to estimate the effect of markers in the estimation population. The number 17 was chosen aiming at providing sufficient number of replicates of the analyses while keeping also an adequate sample size for cross-validation. The groups were formed randomly.

All of the estimated marker effects were applied to the validation population to predict the genomic value of the individuals. Their marker incidence matrix (Z_v), which corresponds to the marker genotypes for the validation population, was multiplied by the estimated effects of each marker and summed the general estimated average. Because the phenotypic value is known when validating the results, it is possible to evaluate the correlation of the genetic value predicted by the phenotype observed in all individuals. This correlation is known as the predictive ability ($r_{y\hat{y}}$) of genomic selection for estimating the phenotypes, and it is theoretically determined by selection accuracy ($r_{g\hat{g}}$) multiplied by the square root of individual heritability (h), i.e., $r_{y\hat{y}} = r_{g\hat{g}}h$ and $r_{gy\hat{g}} = r_{y\hat{y}}/h$ (Resende 2008). The heritability values were estimated from this analysis through estimation of variance component by REML.

Comparisons between phenotypic selection and genomic selection

We compared the genomic selection with the phenotypic selection for selection gain per unit time. For this, the factors used to calculate the selection gain when comparing genomic selection with phenotypic selection were selection accuracy and the possible reduction in time for selecting individuals.

For the comparison effect, the maximum accuracy value likely to be used based on phenotypic data through traditional BLUP selection using pedigree information. That was obtained by analyzing phenotypic data via REML/BLUP using the Selegen-REML/BLUP software. These accuracies were then compared with those obtained through analysis via genomic selection. The relationship was evaluated taking into account the expected reduced generation time in different numbers of years using early selection with only genotypic data.

Results and discussion

Experimental field data

In general, the experiment was satisfactory conducted with normal plant growth and densities. The average SW was 31.87 t ha^{-1}, ranged from 7.32 to 69.46 t ha^{-1} (Table 1). This result agreed with range of 33.1–42.6 t ha^{-1} obtained by Kamau et al. (2010), when analyzing segregant populations from crosses between Kenya landraces versus varieties from IITA (International Institute of Tropical Agriculture). FRY mean was 24.47 t ha^{-1}, with range of 0.75–79.03 t ha^{-1}, while DMC mean was 35.61 %, with range of 21.97–41.51 %. These results were in agreement with Ojulong et al. (2008) analyzing 627 genotypes of a clonal evaluation trial, whose estimates of FRY ranged

Table 1 Simple statistics of agronomic variables evaluated on the genotypes

Variable	Average	Standard deviation	Minimum	Maximum	CV
SW	31.87	15.21	7.32	69.46	27.72
FRY	24.47	17.09	0.75	79.03	34.26
AC	23.29	6.44	20.72	25.74	8.86
DMC	35.61	6.17	21.97	41.51	5.79
S-Y	8.71	5.38	2.68	23.10	33.87

SW shoot weight, *FRY* fresh root yield, *AC* amylose content, *DMC* dry matter content, *S-Y* starch yield, *CV* coefficient of variation

from 9.64 to 63.66 t ha^{-1} and DMC from 20.65 to 41.25 %. Mean FRY was 34.37 t ha^{-1} and 33.47 % for DMC.

For AC, genotypes presented 23.29 % in average and a range of 20.72–25.74 %. In other works, average AC in a sample of more than 4,000 different cassava genotypes was reported to be 20.7 %, ranging from 15 to 26 % (Sánchez et al. 2009). Mean S-Y was 8.71 t ha^{-1}, with S-Y ranging from 2.68 to 23.10 t ha^{-1}.

Selection accuracy as a function of heritability

Estimates of the predictive ability, selection accuracy, and regression coefficients between the observed and predicted values, as well as the calculation of the degree of predictability for all traits evaluated are displayed in Table 2. These values represent the maximum possible accuracy value in this population based on the analysis of all SNPs tested (390). A broad sense heritability for the level of averages (h^2) for all traits obtained in the REML analysis of phenotypic data is displayed in the last line (bolded) of Table 2. We observed modest estimates of h^2 for AC (0.27) but high estimates for SW (0.67), FRY (0.72), DMC (0.67), and S-Y (0.69).

The estimated h^2 for FRY was lower than that obtained by Asante and Dixon (2002) using 10 genotypes in several African countries (0.54). In relation to DMC, which correlates well with the starch content, reports in the literature indicate variation of h^2 from 0.42 to 0.97 (Kawano et al. 1998; Kizito et al. 2007; Ntawuruhunga and Dixon 2010). However, comparisons between these estimates are difficult to interpret because different studies use different genetic material, harvest and planting seasons, and trait measuring techniques.

Using genomic prediction where h^2 was estimated based on experimental data, we observed low selection accuracy values for all traits evaluated. The regression coefficients involving the predicted genomic value and the phenotype observed were also of low magnitude and very far from a value of 1. Thus, the heritability values that provided an estimated β value close to unity for SW, FRY, AC, DMC, and S-Y were 0.25, 0.25, 0.03, 0.20, and 0.26, respectively (Table 2). These estimates of heritability, which were detected using the 390 SNP markers, are effective at predicting the actual magnitudes of the differences

Table 2 Estimates of predictive ability ($r_{y\hat{y}}$), selection accuracy ($r_{g\hat{g}}$), regression coefficient among the observed and predicted values (β) and coefficient of determination (R^2) for SW, FRY, AC, DMC and S-Y depending on the heritability (h^2) used in the prediction for 390 SNP markers

SW					FRY					AC					DMC					S-Y				
h^2	$r_{y\hat{y}}$	$r_{g\hat{g}}$	R^2	β	h^2	$r_{y\hat{y}}$	$r_{g\hat{g}}$	R^2	β	h^2	$r_{y\hat{y}}$	$r_{g\hat{g}}$	R^2	β	h^2	$r_{y\hat{y}}$	$r_{g\hat{g}}$	R^2	β	h^2	$r_{y\hat{y}}$	$r_{g\hat{g}}$	R^2	β
0.25	0.34	0.68	0.46	1.00	0.25	0.31	0.61	0.37	0.99	0.03	0.05	0.27	0.08	0.91	0.20	0.22	0.50	0.25	0.99	0.26	0.32	0.63	0.40	0.99
0.30	0.34	0.62	0.38	0.93	0.30	0.30	0.55	0.31	0.91	0.05	0.06	0.27	0.08	0.80	0.30	0.22	0.40	0.16	0.77	0.30	0.32	0.58	0.34	0.93
0.40	0.33	0.52	0.27	0.82	0.40	0.30	0.47	0.22	0.78	0.10	0.07	0.23	0.05	0.60	0.40	0.21	0.34	0.11	0.62	0.40	0.31	0.49	0.24	0.80
0.67	**0.31**	**0.37**	**0.14**	**0.58**	**0.72**	**0.26**	**0.31**	**0.10**	**0.49**	**0.27**	**0.09**	**0.17**	**0.03**	**0.38**	**0.67**	**0.17**	**0.20**	**0.04**	**0.33**	**0.69**	**0.27**	**0.33**	**0.11**	**0.52**

Numbers in bold refer to the broad sense heritability (h^2) for all traits, obtained by REML analysis of phenotypic data

between the individuals evaluated. Because all traits were evaluated in the complete set of SNP markers, the results displayed in Table 2 indicate the highest accuracy obtained for S-Y compared to other traits. Lower accuracies were observed for AC (0.27). These observations contribute to evaluating the possibilities for successful selection.

In plant species, Heffner et al. (2010) evaluated the cost of genomic selection in relation to MAS and indicated that 0.5 selection accuracy would ensure twice the gain per year compared to MAS in low-investment wheat improvement programs. Furthermore, this selection accuracy would triple the gain in high-investment corn improvement programs. In addition, in one of the first simulated studies of genomic selection, Meuwissen et al. (2001) reported gains in relation to MAS using different methods. They reported gains of 130 % using RR-BLUP, 151 % using Bayes A, and 167 % using Bayes B.

Analysis of markers associated with phenotypes

The markers that had their effects estimated from the phenotypic data were selected after analyzing the association using simple marker regression and association using an F statistic at 5 % significance level. Out of all the markers analyzed, we observed significant association for 118, 92, 56, and 97 SNPs for the traits SW, FRY, DMC, and S-Y, respectively. Because only a 0.03 heritability was detected using the 390 SNPs for the AC trait, this phenotype was removed from the association analysis and other assessments.

Simulations of genomic selection in corn demonstrate that using varying numbers of markers (i.e., 128, 256, 512, or 768) did not lead to significant differences in genomic gains with heritability equal to 0.20 or 20 QTLs that governed the trait. However, with higher heritability (0.50–0.80) or higher numbers of QTLs (40–100), the responses were improved using at least 256 SNPs (Bernardo and Yu 2007).

There was a relationship between the number of significant markers and the magnitude of heritability recovered by the markers ($r = 0.98$). This likely occurred because when analyzing SNPs, the possibility exists that some genomic regions coding for some traits, such as SW, may be more concentrated with SNPs. However, this hypothesis cannot be confirmed because these loci are not yet mapped in the cassava genome. Furthermore, genomic selection differs from other molecular improvement strategies, such as associative mapping and linkage, because mapping the effect of individual genes is not the objective. Instead, the objective is to obtain an efficient estimate of the genetic values as a function of a molecular marker series that ideally covers the species' entire genome (Jannink et al. 2010).

In major crops, such as corn, whose availability of panels for analyzing large numbers of SNP markers is already a reality, GWAS have presented surprising results. Weng et al. (2011) identified a total of 204 SNPs covering 105 genomic coding loci related to plant height by analyzing 284 corn strains with more than 41,000 uniformly spaced SNPs throughout the species' genome. Out of the total number of SNPs, four loci were associated with genes from the biosynthetic pathways for gibberellin (GA) and auxin, as well as epigenetic pathways that may be involved in natural variation leading to a dwarf phenotype in corn strains, which until now remained undetected using other strategies.

Studies have shown that the genomic association involved in populations of specially constructed mapping is effective for revealing the genetic basis of important agronomic traits. Tian et al. (2011) demonstrated that the genetic architecture of corn leaf traits is dominated by small effects, with both epistasis and pleiotropy, particularly related to variations in *liguleless* genes that contribute to leaves being more erect.

For cassava, these results are preliminary due to the low number of SNP markers evaluated. There is still a need to generate high-quality genomic sequences aligned to a reference genome aimed at producing a high number of informative markers for detecting more markers associated with the phenotypes analyzed or other phenotypes of interest. However, our results indicate significant associations between SNPs and phenotypes that should be considered in genetic studies of this species.

Cross-validation using only informative SNPs

For the cross-validation analysis, we used 21 individuals for validation and 336 individuals in the estimation population using the Jackknife resampling strategy. We used only informative markers with evidence of association. The prediction equations of genetic/genomic values were tested for their accuracies in this independent sample. In this case, the

genetic/genomic values are predicted and subjected to correlation analysis with the observed phenotypic values. Because the validation sample was not involved in predicting the effect of the markers, the errors of the genetic/genomic values and phenotypic values are independent; thus, the correlation between these values is predominantly genetic in nature and is equivalent to predictive ability (Resende et al. 2010). Table 3 displays the results of the correlation between the predicted genomic values and phenotypes of individuals in the validation population associated with estimated heritability and explained only by informative markers.

Interestingly, using only the 118 most informative SNPs, we detected the same heritability as when we used the total set of markers for the SW trait (0.25) with optimal regression fitted ($\beta = 1.0$). However, estimated accuracy increased from 0.68 (Table 2) to 0.83 (Table 3), which indicates a dramatic improvement in the prediction of the individual's phenotype.

In contrast, the heritability detected by informative markers for the traits FRY, DMC, and S-Y were lower than broad sense heritability from phenotypic data, approximately 24 %, 50 %, and 23 %, respectively (Table 3). However, there was higher predictive ability than when using all 390 markers (Table 2). This reflects higher internal and intrinsic coherence of the data for providing information on the predicted phenotype when the estimation used larger sample sizes per marker (i.e., higher ratio of N individuals/n markers). This approach of using fewer markers was named the RR-BLUP B method by Resende et al. (2012a).

Table 3 Predictive ability and accuracy of genomic selection using only the most informative markers for each trait, such as SW, FRY, DMC and S-Y

Parameters	SW	FRY	DMC	S-Y
NIM	118	92	56	97
h^2	0.25	0.19	0.10	0.20
$r_{y\hat{y}}$	0.42	0.38	0.30	0.39
$r_{g\hat{g}}$	0.83	0.76	0.67	0.77
R^2	0.69	0.58	0.45	0.59
β	1.00	1.00	0.98	1.00

NIM number of informative markers, $r_{y\hat{y}}$ predictive ability, $r_{g\hat{g}}$ accuracy of genomic selection, β regression coefficient among the observed and predicted values, R^2 coefficient of determination

For the FRY trait, there was a 24.6 % increase in the accuracy of the cross-validation based on using only the 92 most informative SNPs. For the DMC trait, there was a 34 % increase in accuracy compared to an $r_{g\hat{g}}$ of 0.50 in the analysis using all SNPs and a heritability of 0.20 (Tables 2, 3). In addition, for this trait, despite the low heritability detected using genomic selection ($h^2 = 0.10$) with 56 SNPs that are significantly associated with the phenotype, it is possible to obtain high selection accuracy (0.67) due to higher predictive ability.

Following the same trend as the other traits, increased selection accuracy ($r_{g\hat{g}}$ from 0.63 to 0.77) was observed for S-Y when using only 97 SNPs. When evaluating the use of genomic selection in a *Pinus taeda* population, Resende et al. (2012b) obtained prediction accuracies varying from 0.65 to 0.75 and from 0.63 to 0.74 for plant diameter and height, respectively. Thus, the selection accuracy values calculated for yield and quality traits in cassava roots appear to be consistent with those determined for other crops using the RR-BLUP method. Guo et al. (2012) analyzed 25 corn populations and used the RR-BLUP method to obtain accuracy averages of 0.43, 0.43, and 0.25 for days to silking, days to anthesis, and the interval between anthesis and silking, respectively.

In cassava, compared to the analysis based on a previous selection using only informative markers, the selection accuracies were lower than when using all genotyped markers for the majority of the traits. According to Resende et al. (2012b, c), the reason for this inferiority may be because many of the QTLs explain a very small fraction of genetic variation, and in this case, even when markers linked to these QTLs are obtained, the error in estimating the effects is higher than the gain that these effects could provide for the accuracy performed.

Although genomic selection is a method for predicting improvement values based on using a large number of molecular markers, if many markers have zero effect, and the effects are estimated to be different from zero, then their cumulative effects increase noise in the estimates (Goddard and Hayes 2007; Schulz-Streeck et al. 2011). Molecular markers are more useful for genomic selection for higher linkage disequilibrium with the QTLs, and thus, an initial selection of more informative markers for predicting genetic values may be interesting, as was also reported

by other authors (Hayes et al. 2009a; Macciotta et al. 2009; Schulz-Streeck et al. 2011). However, it is necessary to assess the adoption of this pre-selection with criteria because the increased accuracy may not be the same for different crops and assessments, which may in many cases promote significant losses in selection accuracy (Weigel et al. 2009).

Other authors also observed that predictive ability decreases with increased numbers of markers in the model. Indeed, there is a linear increase in the accuracy of genomic selection using the RR-BLUP method (Fernando et al. 2007; Resende et al. 2010, 2012a; Cavalcanti and Resende 2012). In addition, Zhang et al. (2010) demonstrated that the selection accuracy also decreases with advancing selection generations. In a simulated study, reductions in accuracy per generation varied from 0.021 to 0.036 depending on the method used. According to Guo et al. (2012), in corn, densities of markers corresponding to an average distance of 10 cM (approximately 156 markers) were sufficient to detect the linkage disequilibrium between the QTLs and the markers in the genomic selection of populations derived from biparental crosses. Above this density, improved accuracies tended to be zero, which indicates that the 10 cM interval was sufficient to capture the linkage disequilibrium between the QTLs and markers.

Nevertheless, in relation to the number of markers, Crossa et al. (2010) demonstrated that from a modest number of individuals (between 264 and 284) and molecular markers (between 1,135 and 1,148), especially for major crops such as corn, models of genomic selection can achieve high predictive ability of the genetic values for traits of economic interest under contrasting environmental conditions. The results of Zhang et al. (2011) corroborate these findings because there is little loss of accuracy when using fewer markers (100 compared to 10,031). Thus, we believe that in cassava, even with low numbers of markers (i.e., 118, 92, 56, and 97), informative SNPs were sufficient to maximize the accuracy in the validation population for SW, FRY, DMC, and S-Y, respectively. In addition, it also provided significant improvement in selection reliability, with R^2 values of 0.69, 0.58, 0.45, and 0.59 for SW, FRY, DMC, and S-Y, respectively. Solberg et al. (2008) used data from microsatellite markers and SNPs to analyze how the accuracy of the predicted genetic values responded to changes in the marker densities used for genomic selection.

Training population

In the present study, the training population was composed of several cassava genotypes consisting of cultivars, local varieties, and clones in the validation phase, which represent a genetic background with high genetic variability. Thus, this training population has a high concentration of favorable alleles for cassava yield and root quality, which allows gains using genomic selection over several years.

To illustrate these observations on the training population, we used the BLUP process to obtain estimates of genetic gain for all traits based on the predicted genetic value of the individuals that resulted from the predicted genetic effects over the general average of the phenotypes. The genetic gain was obtained by selecting approximately 28 % of the best genotypes for each of the SW, FRY, DMC, and S-Y traits (Table 4). Selection based on genetic values for the SW trait allows us to obtain gains of 31.76 %. Gains are even higher for FRY (39.75 %), increasing the average yield of the selected population to more than 40 t ha^{-1}, which is much higher than the Brazilian national average of approximately 13.86 t ha^{-1} (IBGE 2012). Gains above 37 % can also be obtained using S-Y, which is highly interesting for the industry because higher starch percentages provide higher industrial yield. In contrast, the possible gains for DMC are modest (7.61 %). These gains illustrate the potential of the training population evaluated in this study if effective selection strategies are implemented.

According to Rutkoski et al. (2011), the ideal situation would be to generate a new training population for each family derived from biparental crosses. This would lead to high selection accuracy because the allele frequencies, the effects of QTLs, and the genetic

Table 4 Selection gains (SG) for SW, FRY, DMC and S-Y, using individual BLUP over phenotypic data

Trait	$\bar{X}_p$	$\bar{X}_m$	SG (%)
SW	31.88	46.72	31.76
FRY	24.47	40.62	39.75
DMC	35.61	38.54	7.61
S-Y	8.71	14.04	37.94

$\bar{X}_p$ mean of all accessions tested, $\bar{X}_m$ mean of the 100 best genotypes for each trait, *SG* selection gain in %

backgrounds must be similar among the candidates of both the selection and training populations. However, it is necessary for the training population of each cross to have a phenotype in different environments of interest prior to the genomic selection modeling. In contrast, using a training population representative of the improvement program, the cassava training population example may contribute to increasing the accuracy of estimating each individual's genome in addition to reducing the selection cycles.

According to Rutkoski et al. (2011), introducing genetic diversity into improvement programs is of paramount importance to improve traits; for example, in disease resistance where important alleles may not be present in selected genotypes. Therefore, incorporating new sources of genetic diversity, such as using local varieties or exotic germplasm, is a major challenge for improvement programs using genomic selection. Furthermore, simulations in cattle used to assess the potential use of training populations consisting of several different breeds to calculate genomic values of a particular breed indicate high accuracies when using high-density markers (de Roos et al. 2009; Hayes et al. 2009a). Thus, training populations with high genetic diversity representative of improvement programs may be an interesting strategy to obtain superior selective accuracies in different families derived from these training populations. Therefore, the strategy implemented in this training population of cassava has a great agronomic potential to ensure genetic gains using different backgrounds.

Selection gains with genomic selection

The estimated gains were used to compare the efficiency of genomic selection with phenotypic selection, which has been traditionally used in cassava improvement programs. Usually a selection cycle in cassava takes at least 4 years. This cycle includes phenotypic assessment in half- or full-sib progenies for 1 year at each of the following stages: seedling assessment, clonal test, preliminary test, advanced test, and then intercrossing of the best individuals. Thus, because one of the applications of genomic selection is to promote early selection, the gain was assessed per unit time and took into account the reduced selection cycle for 3, 2, and 1 years (Table 5).

The accuracy values obtained via genomic selection are lower than the maximum values obtained in

Table 5 Superiority of genomic selection in relation to maximum accuracy obtainable via phenotypic selection due to the reduction of generation time in years from phenotypic (PB) and genomic breeding (GB), for SW, FRY, DMC and S-Y

PB	GB	SW				FR				DMC				S-Y			
		PA	GA	EFI	SUP	PA	GA	EFI	SUP	PA	GA	EFI	SUP	PA	GA	EFI	SUP
4	4	0.96	0.84	86.98	−13.02	0.97	0.76	78.45	−21.55	0.96	0.67	69.71	−30.29	0.96	0.77	79.82	−20.18
4	3	0.96	0.84	115.97	15.97	0.97	0.76	104.60	4.60	0.96	0.67	92.95	−7.05	0.96	0.77	106.43	6.43
4	2	0.96	0.84	173.96	73.96	0.97	0.76	156.90	56.90	0.96	0.67	139.42	39.42	0.96	0.77	159.64	59.64
4	1	0.96	0.84	347.92	247.92	0.97	0.76	313.80	213.80	0.96	0.67	278.85	178.85	0.96	0.77	319.28	219.28

PA phenotypic accuracy, *GA* genomic accuracy, *EFI* efficiency = (GA·PB)/(PA·GB), *SUP* superiority of genomic selection in %

the phenotypic selection for all traits analyzed and accounted for the set of markers significantly associated with each trait and the same 4 years required for both conventional (phenotypic) and genomic selection (Table 5). In this situation, traits with low genomic accuracy, such as DMC, exhibited approximately 30 % inferiority compared to phenotypic selection. However, the high impact of using genomic selection would be to reduce the improvement cycles to improve early selection of the best individuals. Reducing the selection cycle of the best individuals by 3 years would result in higher genetic selections for SW, FRY, and S-Y of 15.97 %, 4.60 %, and 6.43 %, respectively. However, even under these conditions, the genomic selection is phenotypically less for DMC (−7.05 %).

When reducing the selection cycle by half, the selection gains may vary from 73.96 %, 56.90 %, 39.42 %, and 59.64 % for SW, FRY, DMC, and S-Y, respectively. Reducing the selection cycle to 1 year would only take into consideration the seedling assessments and the generation of new crosses. In this case, the gains would be even higher, theoretically reaching 178.85 % for DMC and 247.92 % for SW (Table 5). However, this situation is very optimistic and difficult to achieve in practice due to the limitations of genotyping and the generation of segregating populations in most improvement programs. Therefore, we believe that reducing the selection cycle in cassava by half, the currently practiced techniques used to test the actual value of a genotype would be the most realistic situation possible. Under this condition, the theoretical gains possible using genomic selection vary from 39.42 %, 56.90 %, and 73.96 % for DMC, FRY, and SW, respectively. These results would bring an enormous increase to the efficiency of cassava improvement programs.

In practice, incorporating genomic selection could dramatically reduce the time required for completion of a cycle of genetic improvement by eliminating the testing phase. Genomic selection of cassava could start with a large population of seedling plants coming from botanical seed. At this stage, genome selection would be applied and the stem cuttings from selected genotypes could then be taken and sent back to the crossing block to initiate a new cycle of selection. After three cycles of genomic selection, some stem cuttings should be planted for field evaluation and validation of the selection made through genome selection. In this scenario, rapid pyramiding of favorable alleles could be made by crossing the best allelic complement across QTL throughout the genome. As a consequence, the breeding phase of generation of hybrids may be completed in 3 years. In parallel, selected seedlings can be clonally replicated and established in clonal trials to verify their performance relative to elite material in the second phase. Nevertheless, field testing has remained the most time-consuming phase in genetic improvement of cassava, lasting typically for 6–7 years from clonal to advanced trails. In addition there is a bottleneck related to the limited production of planting material in the initial steps. However, the second stage can be reduced to 4–5 years by implementing rapid multiplication of stem cuttings.

Depending on the assessment environment, by applying a 50 % reduced selection cycle, the efficiency of genomic selection per unit time increased from 53 to 92 % for plant diameter and from 58 to 112 % for height compared to the phenotypic selection in *P. taeda* (Resende et al. 2012a. The selective accuracies remained high only in assessment environments belonging to the same environmental stratification, and the accuracies varied in different assessment environments, which resulted in different selection gains for the same trait. In contrast, considering that the earlier phenotypic predictions of individuals results in the increased chance of adopting genomic selection by allowing higher gain per unit time, we assessed how models developed at an early age (1 and 2 years) could contribute to the phenotype prediction at older ages (6 years). The results indicated that the prediction models obtained in the training populations at early ages did not exhibit satisfactory accuracy to predict genetic values at more advanced stages in *P. taeda* populations.

The major advantage of genomic selection is the ability to predict the potential and likelihood of reducing the generation interval with high accuracy by estimating precise genetic values at early selection stages, including even before birth for animals (Meuwissen et al. 2001; Muir 2007; Hayes et al. 2009b). However, according to Resende et al. (2012a), it is necessary to promote training population assessment at an adequate age to obtain a higher selection accuracy in addition to each environmental stratum from which the effect of markers were estimated because the genotype–environment interaction drastically affects the transferability of models from one edaphoclimatic region to another.

Recently, Bernardo and Yu (2007) demonstrated that genomic selection can overcome inclusive selection strategies that already include marker use in improvement, e.g., the use of marker-assisted recurrent selection (MARS) in corn. In all simulations, genomic selection was higher than MARS; however, considering the maximum responses, we suggest that genomic selection is most useful for complex traits controlled by many QTLs with low heritability.

Currently, most cassava improvement programs do not use MAS in practice. This is possibly due to the high cost of genotyping or low selection efficiency based on QTLs with small effects and generally obtained in small populations with a restricted genetic basis. However, the discovery of a high number of SNPs, the rapid development of automated genotyping technologies, and even DNA extraction may easily overcome the first difficulty. Despite the need to generate large-scale SNP genotyping panels in cassava, the results of this pilot trial indicate that genomic selection could revolutionize selection strategies in cassava improvement programs, considering the possibility of high genetic gains in the short term, even using low numbers of SNP markers. However, the advantages of genomic selection identified in this study should not be expected for any trait and in any crop. To validate the efficiency of genomic selection and to obtain a broader understanding, further studies are needed to validate the models for other traits and assessment environments.

Several studies with simulated data have reported important genetic gains in plant improvement using genomic selection (Bernardo and Yu 2007; Wong and Bernardo 2008; Mayor and Bernardo 2009; Zhong et al. 2009). In dairy cattle, Schaeffer (2006) indicated that saving time and cost using genomic selection with a 0.75 accuracy could double the selection gain and provide up to 92 % savings compared to traditional selection methods. However, there are still many hypothesized uncertainties about effective gain in real situations. Nevertheless, studies of genomic selection implemented based on data from field experiments with cassava indicate the possibility of successful use of this method to optimize the selection of important traits in cassava, whose phenotype involves high costs of assessing experiments over several years and different locations and is highly subject to harsh natural conditions. In these cases, it is possible to bypass the high cost of testing and evaluating the progeny, and consequently, the time required to complete an improvement cycle. Even with reduced accuracy in genomic selection over selection generations, monitoring the accuracy of prediction models and their constant recalibration may allow for the use of this technique effectively in various plant species, as has been performed in animals.

Recently, Heffner et al. (2009) concluded that the high correlation between the genetic phenotypic values and the predicted genetic/genomic values are already sufficient. This suggests that we consider basing plant selection improvement programs only on molecular markers. According to Wong and Bernardo (2008), another argument in favor of genomic selection is that genotyping costs will further decrease in the future, while phenotyping costs do not exhibit the same downward trend because they always depend on human resources and agricultural inputs whose price history is increasing.

Acknowledgments The authors thank CNPq, CAPES and FAPESB for fellowships and financial support. E.J.O. and M.D.V.R. are fellow CNPq researchers.

References

Asante IK, Dixon AGO (2002) Analysis of phenotypic stability in ten cassava genotypes in three West African countries. West Afr J Appl Ecol 3:43–48

Bernardo R, Yu J (2007) Prospects for genome-wide selection for quantitative traits in maize. Crop Sci 47:1082–1090

Cavalcanti JJ, Resende MDV (2012) Predição simultânea de efeitos de marcadores e seleção genômica ampla em cajueiro. Rev Bras Frutic (in press)

Ceballos H, Iglesias CA, Perez JC, Dixon AGO (2004) Cassava breeding: opportunities and challenges. Plant Mol Biol 56:503–516

CIAT (2008) Annual report from SBA-2 Project. Improved cassava for the developing world. CIAT, Cali

Cortez DF, Reilly K, Okogbenin E, Beeching JR, Iglesias C, Tohme J (2002) Mapping wound-response genes involved in post-harvest physiological deterioration (PPD) of cassava (*Manihot esculenta* Crantz). Euphytica 128:47–53

Crossa J, Campos G, Pérez P, Gianola D, Burgueño J, Araus JL, Makumbi D, Singh RP, Dreisigacker S, Yan J, Arief V, Banziger M, Braun H-J (2010) Prediction of genetic values of quantitative traits in Plant Breed using pedigree and molecular markers. Genetics 186:713–724

Daetwyler HD, Pong-Wong R, Villanueva B, Woolliams JA (2010) The impact of genetic architecture on genome-wide evaluation methods. Genetics 185:1021–1031

de Roos APW, Hayes BJ, Goddard ME (2009) Reliability of genomic predictions across multiple populations. Genetics 183:1545–1553

Dekkers JCM (2004) Commercial application of marker- and gene-assisted selection in livestock: strategies and lessons. J Anim Sci 82:313–328

Dekkers JCM (2007) Marker-assisted selection for commercial crossbred performance. J Anim Sci 85:2104–2114

FAO (2011) Disponível em, p. http://www.faostatfaoorg/defaultaspx. Accessed 11 Aug 2011

Fernando RL, Habier D, Stricker C, Dekkers JCM, Tottir LR (2007) Genomic selection. Acta Agric Scand A 57:192–195

Gianola D, de los Campos G, Hill WG, Manfredi E, Fernando R (2009) Additive genetic variability and the Bayesian alphabet. Genetics 183:347–363

Goddard ME, Hayes BJ (2007) Genomic selection. J Anim Breed Genet 124:323–330

Gomes JC, Silva J (2006) Correção da acidez e adubação. In: Souza LS, Farias ARN, Mattos PLP de, Fukuda WMG (eds) Aspectos socioeconômicos e agronômicos da mandioca. Embrapa Mandioca e Fruticultura Tropical, Cruz das Almas, pp 215–247

Guo Z, Tucker DM, Lu J, Kishore V, Gay D (2012) Evaluation of genome-wide selection efficiency in maize nested association mapping populations. Theor Appl Genet 124:261–275

Hayes BJ, Bowman PJ, Chamberlain AJ, Goddard ME (2009a) Invited review: genomic selection in dairy cattle: progress and challenges. J Dairy Sci 92:433–443

Hayes B, Bowman P, Chamberlain A, Verbyla K, Goddard M (2009b) Accuracy of genomic breeding values in multi-breed dairy cattle populations. Genet Sel Evol 41:51. doi: 10.1186/1297-9686-41-51

Heffner EL, Sorrells ME, Jannink JL (2009) Genomic selection for crop improvement. Crop Sci 49:1–12

Heffner EL, Lorenz AJ, Jannink JL, Sorrells ME (2010) Plant Breed with genomic selection: potential gain per unit time and cost. Crop Sci 50:1681–1690

IBGE (2012) Produção agrícola municipal Disponível em. http://www.sidraibgegobr/bda/. Accessed 26 Jan 2012

Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in Plant Breed: from theory to practice. Brief Funct Genomics 9:166–177

Jenkins S, Gibson N (2002) High-throughput SNP genotyping. Comp Funct Genomics 3:57–66

Jorge V (2000) Cartographie de la résistance du manioc à la bacteriose vasculaire du manioc causée par *Xanthomonas axonopodis* pv. *manihotis*. PhD Thesis, Université Paris XI, France

Jorge V, Fregene M, Duque MC, Bonierbale MW, Tohme J, Verdier V (2000) Genetic mapping of resistance to bacterial blight disease in cassava (*Manihot esculenta* Crantz). Theor Appl Genet 101:865–872

Jorge V, Fregene M, Velez CM, Duque MC, Tohme J, Verdier V (2001) QTL analysis of field resistance to *Xanthomonas axonopodis* pv. *manihotis* in cassava. Theor Appl Genet 102:564–571

Kamau J, Melis R, Laing M, Derera J, Shanahan P, Ngugi E (2010) Combining the yield ability and secondary traits of selected cassava genotypes in the semi-arid areas of Eastern Kenya. J Plant Breed Crop Sci 2:181–191

Kawano K, Narintaraporn K, Narintaraporn P, Sarakarn S, Limsila A, Limsila J, Suparhan D, Sarawat V, Watananonta W (1998) Yield improvement in a multistage breeding program for cassava. Crop Sci 38:325–332

Kawuki RS, Ferguson M, Labuschagne M, Herselman L, Kim D-J (2009) Identification characterisation and application of single nucleotide polymorphisms for diversity assessment in cassava (*Manihot esculenta* Crantz). Mol Breed 23:669–684

Kizito EB, Ronnberg-Wastljung AC, Egwang T, Gullberg U, Fregene M, Westerbergh A (2007) Quantitative trait loci controlling cyanogenic glucoside and dry matter content in cassava (*Manihot esculenta* Crantz) roots. Hereditas 144: 129–136

Kunkeaw S, Yoocha T, Sraphet S, Boonchanawiwat A, Boonseng O, Lightfoot DA, Triwitayakorn K, Tangphatsornruang S (2011) Construction of a genetic linkage map using simple sequence repeat markers from expressed sequence tags for cassava (*Manihot esculenta* Crantz). Mol Breed 27:67–75

Long N, Gianola D, Rosa GJM, Weigel KA, Avendano S (2007) Machine learning classification procedure for selecting SNPs in genomic selection: application to early mortality in broilers. J Anim Breed Genet 124:377–389

Lopez C, Piegu B, Cooke R, Delseny M, Tohme J, Verdier V (2005) Using cDNA and genomic sequences as tools to develop SNP strategies in cassava (*Manihot esculenta* Crantz). Theor Appl Genet 110:425–431

Lopez CE, Quesada-Ocampo LM, Bohorquez A, Duque MC, Vargas J, Tohme J, Verdier V (2007) Mapping EST-derived SSRs and ESTs involved in resistance to bacterial blight in *Manihot esculenta*. Genome 50:1078–1088

Macciotta NPP, Gaspa G, Steri R, Pieramati C, Carnier P, Dimauro C (2009) Pre-selection of most significant SNPS for the estimation of genomic breeding values. BMC Proc 3(Suppl 1):S14

Mayor PJ, Bernardo R (2009) Genome-wide selection and marker-assisted recurrent selection in double haploid versus F_2 population. Crop Sci 49:1719–1725

Meuwissen THE, Goddard ME (2010) Accurate prediction of genetic values for complex traits by whole-genome resequencing. Genetics 185:623–631

Meuwissen THE, Hayes BJ, Goddard ME (2001) Prediction of total genetic value using genome-wide dense marker maps. Genetics 157:1819–1829

Muir WM (2007) Comparison of genomic and traditional BLUP-estimated breeding value accuracy and selection response under alternative trait and genomic parameters. J Anim Breed Genet 124:342–355

Nassar NMA, Ortiz R (2007) Cassava improvement: challenges and impacts. J Agric Sci 145:163–171

Ntawuruhunga P, Dixon AGO (2010) Quantitative variation and interrelationship between factors influencing cassava yield. J Appl Biosci 26:1594–1602

Ojulong H, Labuschangne MT, Fregene M, Herselman L (2008) A cassava clonal evaluation trial based on a new cassava breeding scheme. Euphytica 160:119–129

Okogbenin E, Fregene M (2002) Genetic and QTL mapping of early root bulking in an F_1 mapping population of non-inbred parents in cassava (*Manihot esculenta* Crantz). Theor Appl Genet 106:58–66

Okogbenin E, Fregene M (2003) Genetic mapping of QTLs affecting productivity and plant architecture in a full-sib cross from non-inbred parents in cassava (*Manihot esculenta* Crantz). Theor Appl Genet 107:1452–1462

Olsen KM (2004) SNPs SSRs and inferences on cassava's origin. Plant Mol Biol 56:517–526

Resende MDV (2007) Seleção genômica ampla (GWS) e modelos lineares mistos. In: Resende MDV. Matemática e estatística na análise de experimentos e no melhoramento genético, 1st edn. Embrapa Florestas, Colombo, pp 517–534

Resende MDV (2008) Genômica quantitativa e seleção no melhoramento de plantas perenes e animais. Embrapa Florestas, Colombo

Resende MDV, Lopes S, Silva RL, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. Pesq Flor Bras 56:63–78

Resende MDV, Resende MFR Jr, Aguiar AM, Abad JIM, Missiaggia AA, Sansaloni C, Petroli C, Grattapaglia D (2010) Computação da seleção genômica ampla (GWS). Embrapa Florestas, Colombo

Resende MFR Jr, Muñoz M, Resende MDV, Garrick DJ, Fernando RL, Davis JM, Jokela EJ, Martin TA, Peter GF, Kirst M (2012a) Accuracy of genomic selection methods in a standard dataset of loblolly pine (*Pinus taeda* L.). Genetics 190:1503–1510

Resende MFR Jr, Munõz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M (2012b) Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. New Phytol 193:617–624

Resende MDV, Resende MFR Jr, Sansaloni CP, Petroli CD, Missiaggia AA, Aguiar AM, Abad JM, Takahashi EK, Rosado AM, Faria DA, Pappas GJ Jr, Kilian A, Grattapaglia D (2012c) Genomic selection for growth and wood quality in *Eucalyptus*: capturing the missing heritability and accelerating breeding for complex traits in forest trees. New Phytol 194:116–128

Rutkoski JE, Heffner EL, Sorrells ME (2011) Genomic selection for durable stem rust resistance in wheat. Euphytica 179: 161–173

Sánchez T, Mafla G, Morante N, Ceballos H, Dufour D, Calle F, Moreno X, Pérez JC, Debouck D (2009) Screening of starch quality traits in cassava (*Manihot esculenta* Crantz). Starch 61:12–19

Schaeffer LR (2006) Strategy for applying genome-wide selection in dairy cattle. J Anim Breed Genet 123:218–223

Schulz-Streeck T, Ogutu JO, Piepho H-P (2011) Pre-selection of markers for genomic selection. BMC Proc 5(Suppl 3):S12

Solberg TR, Sonesson AK, Woolliams JA, Meuwissen THE (2008) Genomic selection using different marker types and densities. J Anim Sci 86:2447–2454

Tian F, Bradbury J, Brown J, Hung H, Sun Q, Flint-Garcia S, Rocheford TR, McMullen MD, Holland JB, Buckler ES (2011) Genome-wide association study of leaf architecture in the maize nested association mapping population. Nat Genet 43:159–162

Villanueva B, Fernández J, García-Cortés LA, Varona L, Daetwyler HD, Toro MA (2011) Accuracy of genome-wide evaluation for disease resistance in aquaculture breeding programs. J Anim Sci 89:3433–3442

Weigel KA, de los Campos G, González-Recio O, Naya H, Wu XL, Long N, Rosa GJ, Gianola D (2009) Predictive ability of direct genomic values for lifetime net merit of Holstein sires using selected subsets of single nucleotide polymorphism markers. J Dairy Sci 92:5248–5257

Welsch R, Arango J, Bar C, Salazar B, Al-Babili S, Beltrán J, Chavarriaga P, Ceballos H, Tohme J, Beyer P (2010) Provitamin A accumulation in cassava (*Manihot esculenta*) roots driven by a single nucleotide polymorphism in a phytoene synthase gene. Plant Cell 22:3348–3356

Weng J, Xie C, Hao Z, Wang J, Liu C, Li M, Zhang D, Bai L, Zhang S, Li X (2011) Genome-wide association study identifies candidate genes that affect plant height in Chinese elite maize (*Zea mays* L.) inbred lines. PLoS ONE 6(12):e29229. doi:10.1371/journal.pone.0029229

Whankaew S, Poopear S, Kanjanawattanawong S, Tangphatsornruang S, Boonseng O, Lightfoot DA, Triwitayakorn K (2011) A genome scan for quantitative trait loci affecting cyanogenic potential of cassava root in an outbred population. BMC Genomics 12:266

Wong CK, Bernardo R (2008) Genomewide selection in oil palm: increasing selection gain per unit time and cost with small populations. Theor Appl Genet 116:815–824

Wydra K, Zinsou V, Jorge V, Verdier V (2004) Identification of pathotypes of *Xanthomonas axonopodis* pv. *manihotis* in Africa and detection of quantitative trait loci and markers for resistance to bacterial blight. Phytopathology 94:1084–1093

Zhang Z, Liu JF, Ding XD, Bijma P, de Koning DJ, Zhang Q (2010) Best linear unbiased prediction of genomic breeding values using trait-specific marker-derived relationship matrix. PLoS ONE 5:9. doi:10.1371/journal.pone.0012648

Zhang Z, Ding XD, Liu JF, de Koning D-J, Zhang Q (2011) Genomic selection for QTL-MAS data using a trait-specific relationship matrix. BMC Proc 5(Suppl 3):S15

Zhong S, Dekkers JCM, Fernando RL, Jannink JL (2009) Factors affecting accuracy from genomic selection in populations derived from multiple inbred lines: a barley case study. Genetics 182:355–364