



Selective efficiency of genome-wide selection in *Coffea canephora* breeding

Emilly Ruas Alkimim¹ · Eveline Teixeira Caixeta² · Tiago Vieira Sousa³ · Marcos Deon Vilela Resende⁴ · Felipe Lopes da Silva⁵ · Ney Sussumu Sakiyama⁵ · Laércio Zambolim⁶

Received: 12 December 2019 / Revised: 29 March 2020 / Accepted: 31 March 2020 / Published online: 20 April 2020
© Springer-Verlag GmbH Germany, part of Springer Nature 2020

Abstract

Over the years, breeding programs have sought efficient strategies to select genotypes with superior performance. Genome-wide selection (GWS) emerged in 2001, aiming to increase efficiency and accelerate selection gain. This technique is considered essential in the breeding of perennial species, such as *Coffea canephora*, mainly due to the potential to increase the gain per unit of time. Thus, this study aimed to apply the GWS principle, evaluate the efficiency of this technique in *C. canephora* population using SNP molecular markers, and evaluate eight main phenotypic traits. A total of 165 genotypes were evaluated, being 51 of varietal group of Conilon, 32 of Robusta, and 82 intervarietal hybrids. Through the sequencing of the RAPiD Genomics company, 18,111 SNP markers were identified, of which 14,429 were used after quality analysis. All traits showed good predictive capacity, except for fruit maturation time, fruit size, and yield per plant. The lower values of genomic heritability found for these traits may justify the low values of predictive capacity obtained. The accuracy values estimated were considered as moderate to high, ranging from 67 to 82%. By shortening the cycle time from 6 to 3 years, GWS provided selective efficiency ranging from 22 to 146%. Results revealed that GWS provides higher gains per unit of time. Therefore, GWS proved to be a useful and promising tool for the breeding of *C. canephora* for accurately predicting the individuals' genotypes, shortening the time required to complete the selection cycle and providing gains in selective efficiency per unit of time.

Keywords Coffee · GBLUP · Predictive capacity · Selective accuracy

Communicated by F. P. Guerra

Electronic supplementary material The online version of this article (<https://doi.org/10.1007/s11295-020-01433-3>) contains supplementary material, which is available to authorized users.

✉ Eveline Teixeira Caixeta
eveline.caixeta@embrapa.br

¹ Universidade Federal do Triângulo Mineiro, Iturama, MG, Brazil

² Empresa Brasileira de Pesquisa Agropecuária - Embrapa Café, Bioagro, BioCafé, Universidade Federal de Viçosa, Vicososa, MG 36570-000, Brazil

³ Instituto Federal Goiano, Urutai, MG, Brazil

⁴ Empresa Brasileira de Pesquisa Agropecuária - Embrapa Café, Vicososa, MG, Brazil

⁵ Departamento de Fitotecnia, Universidade Federal de Viçosa, Vicososa, MG, Brazil

⁶ Bioagro, Universidade Federal de Viçosa, Vicososa, MG, Brazil

Introduction

For a long time, genetic breeding programs selected the best individuals using morphological markers. These markers are influenced by the environment and have low selection gain (Toppa and Jadoski 2013). Technology has improved the use of molecular markers, allowing significant progress in the selection of superior individuals by detecting DNA polymorphism (Sousa et al. 2017; Alkimim et al. 2017). In the early 1990s, marker-assisted selection (MAS) was proposed, based on the existence of linkage disequilibrium between DNA markers and genes of interest (Lande and Thompson 1990). When comparing MAS for oligogenic traits with phenotypic selection, the selective efficiency increases and the time required to carry out selection shortens, among other benefits (Noir et al. 2003; Lopez et al. 2013; Romero et al. 2014; Alkimim et al. 2017). Also, this technique enables preventive breeding by allowing the selection of genotypes carrying genes of interest in regions where the pathogen is absent (Alkimim et al. 2017). However, the MAS technique has

proven to be more efficient for monogenic or oligogenic traits with high heritability (Asins et al. 2012; Kemper and Goddard 2012). Agronomic traits, in general, are governed by several genes, compromising the efficiency of MAS. Therefore, a new selection method, known as genome-wide selection (GWS), was developed (Meuwissen et al. 2001). GWS emphasizes the simultaneous prediction of the genetic effects of hundreds or thousands of markers that densely cover the genome. Thus, all quantitative trait loci (QTL) of a quantitative trait are expected to be in linkage disequilibrium with at least some of the markers (Grattapaglia and Resende 2011; Valente et al. 2016). GWS stands out for promoting high selective accuracy and for not requiring the knowledge of the prior location (maps) of the QTL in the chromosomes (Meuwissen 2007; Jannink et al. 2010; de Almeida et al. 2016).

In the GWS approach, genomic estimated breeding value (GEBV) can be predicted by different statistical methodologies, including GBLUP (genomic best linear unbiased prediction) (VanRaden 2008). In this method, the GEBV values are predicted using the kinship matrix, estimated from the information of molecular markers, known as the genomic kinship matrix (G). The GBLUP prediction uses much more information on parentage than phenotypic selection, which is based on pedigree (through the parentage matrix A). Then genomic heritability and accuracy of genomic selection can sometimes be higher than those parameters from phenotypic selection. And this can be explained by the many more genetic relationship in the G (the genomic relationship matrix) than in A (the genetic relationship matrix based on genealogy). This increase in the amount of information by using the genomic matrix G can, sometimes, lead to better and more precise estimations and predictions. For two populations and its hybrid population, the genetic variance and heritability are defined at the interpopulation level (Bernardo, 2010; Resende, 2015).

The GBLUP is advantageous for its simplicity and for the shorter computational time required (Heslot et al. 2015). This method is mostly recommended for polygenic traits, which are governed by several genes of minor effect (VanRaden 2008). GBLUP is suitable for the analysis of continuous traits or outcomes. For non-normally distributed traits such as those evaluated by a score scale, GBLUP can be used with the technique called generalized linear model. The results may not differ so much from those got by using the standard procedure of linear mixed model. This is in line with theory, which preconizes that the higher the number of score scale classes, the smaller the benefit from using the generalized linear model technique (Sousa et al. 2019).

The molecular markers SNPs (single nucleotide polymorphism), used in GWS studies, stand out for being the most common type of polymorphism in genomes, for the possibility of automation, and for being codominant and biallelic (Resende et al. 2008; Liao and Lee 2010). Recently, the development of next-generation sequencing (NGS) platforms

has facilitated the discovery of SNPs, decreasing data point costs. With the identification of SNP markers widely distributed in the species genome, GWS has become a reality, allowing significant gains for several breeding programs (Goddard and Hayes 2007; Meuwissen 2007; Carvalho and Silva 2010; Resende et al. 2012; Fritsche-Neto et al. 2012; Van Eenennaam et al. 2014; Zhao et al. 2015; Sousa et al. 2019).

Although significant, the number of reports regarding GWS in the genus *Coffea*, even for species of commercial importance, such as *Coffea canephora* and *Coffea arabica*, is still low (Ferrão et al. 2017, 2019; Sousa et al. 2019). In a recent study with two populations of a recurrent selection of *C. canephora*, genotyping by sequencing (GBS) showed good potential to be used in coffee breeding programs (Ferrão et al. 2017). *C. canephora* is characterized as allogamous, diploid ($2n = 2x = 22$), with gametophytic self-incompatibility (Leroy et al. 2005). The species stands out for its rusticity, high yield potential, higher soluble solids content, and genetic resistance to coffee leaf rust, caused by the fungus *Hemileia vastatrix* (Zambolim 2016).

This study aimed to apply the GWS principle and evaluate its efficiency in the prediction of genomic-genetic value and in the shortening of the selective cycle in *C. canephora* population, through the RAPiD Genomics sequencing company, by building specific probes in coding and non-coding regions.

Material and methods

Genetic material

The population consisted of clones of the Conilon and Robusta varietal groups and intervarietal hybrids originated from crosses between these groups. The Conilon genetic material was obtained from the Instituto Capixaba de Pesquisa, Assistência Técnica e Extensão Rural (Incaper), and the Robusta material was obtained from the Centro Agronômico Tropical de Investigación y Enseñanza (CATIE). This population composes the breeding program of the Empresa de Pesquisa Agropecuária de Minas Gerais (Epamig), in partnership with the Universidade Federal de Viçosa (UFV) and the Empresa Brasileira de Pesquisa Agropecuária—Café (Embrapa Café), located in Oratórios/MG and Viçosa/MG.

The Conilon and Robusta varietal groups consisted of 51 and 32 genotypes (Table 1), respectively. Also, 82 intervarietal hybrids were obtained by artificial crosses between five genotypes of the Conilon group (male parents) and five genotypes of the Robusta group (female parents), evaluated in the inter-population partial diallel (Table 2).

For the non-crossed genotypes, information came only from the parents of Conilon and Robusta in the experiment and from their parentage with the crossed parents.

Table 1 *Coffea canephora* genotypes used in the genome-wide selection study

Varietal group	Genotypes			
Conilon	UFV 513	UFV 3628-23	UFV 3628-45	UFV 3629-31
	UFV 3627-20	UFV 3628-24	UFV 3628-46	UFV 3629-34
	UFV 3627-24	UFV 3628-26	UFV 3628-5	UFV 3629-4
	UFV 3627-27	UFV 3628-27	UFV 3629-10	UFV 3630-10
	UFV 3627-29	UFV 3628-28	UFV 3629-11	UFV 3630-2
	UFV 3627-30	UFV 3628-29	UFV 3629-17	UFV 3630-5
	UFV 3627-31	UFV 3628-3	UFV 3629-24	UFV 3630-6
	UFV 3628-1	UFV 3628-32	UFV 3629-25	UFV 3630-7
	UFV 3628-10	UFV 3628-35	UFV 3629-26	UFV 3631-10
	UFV 3628-16	UFV 3628-4	UFV 3629-27	UFV 3631-11
	UFV 3628-17	UFV 3628-40	UFV 3629-28	UFV 3631-13
	UFV 3628-2	UFV 3628-42	UFV 3629-29	UFV 3631-9
	UFV 3628-22	UFV 3628-44	UFV 3629-30	–
	Robusta	UFV 3356-71	UFV 3362-118	UFV 3367-101
UFV 3356-76		UFV 3363-125	UFV 3367-105	UFV 3374-29
UFV 3357-93		UFV 3365-144	UFV 3368-52	UFV 3375-65
UFV 3358-88		UFV 3366-134	UFV 3368-58	UFV 3375-66
UFV 3360-169		UFV 3366-138	UFV 3370-50	UFV 3376-8
UFV 3360-171		UFV 3366-139	UFV 3371-19	UFV 3376-9
UFV 3361-148		UFV 3367-96	UFV 3371-20	UFV 3377-12
UFV 3361-151		UFV 3367-97	UFV 3373-36	UFV 514

Table 2 Intervarietal hybrids and description of their crosses, used in the genome-wide selection study

Hybrids	Crosses
H092-2, H092-8, H092-10, H092-20, H092-22, H092-35	UFV 3366-139 x UFV 513
H093-1, H093-10, H093-12, H093-20	UFV 3366-139 x UFV 3629-11
H094-2, H094-5, H094-12, H094-13, H094-18, H094-30	UFV 3367-98 x UFV 513
H095-9, H095-11, H095-14, H095-25, H095-28	UFV 3366-139 x UFV 3627-31
H097-8, H097-13, H097-16, H097-17, H097-18, H097-19	UFV 3374-28 x UFV 3627-31
H098-1, H098-2, H098-3, H098-5, H098-13	UFV 3365-144 x UFV 3629-11
H099-1, H099-2, H099-7, H099-8, H099-10	UFV 3365-144 x UFV 513
H0910-4, H0910-5, H0910-6, H0910-7, H0910-8	UFV 3367-98 x UFV 3629-11
H0911-2, H0911-3, H0911-4, H0911-5, H0911-6, H0911-7	UFV 3367-98 x UFV 3628-2
H0912-1, H0912-3, H0912-4, H0912-6, H0912-7, H0912-8	UFV 3366-139 x UFV 3628-2
H0913-1, H0913-3, H0913-4, H0913-5, H0913-7	UFV 3365-144 x UFV 3628-2
H0914-2, H0914-3, H0914-4, H0914-5, H0914-6	UFV 3374-28 x UFV 513
H0915-1, H0915-2, H0915-3, H0915-4, H0915-5	UFV 3373-36 x UFV 3628-2
H0916-1, H0916-2, H0916-4	UFV 3367-98 x UFV 3629-25
H0917-1, H0917-2, H0917-3	UFV 3365-144 x UFV 3629-25
H0918-1, H0918-2, H0918-3	UFV 3373-36 x UFV 513
H0919-2	UFV 3373-36 x UFV 3629-11
H0920-1	UFV 3373-36 x UFV 3629-25
H0921-1	UFV 3374-28 x UFV 3629-11
H0922-1	UFV 3373-36 x UFV 3627-31

Phenotypic evaluations

The experiment was established in an incomplete block design with up to 35 replicates and single tree plots. It included hybrids and parents. Phenotypic evaluations were carried out for eight traits during three consecutive years (2014–2016). Five categorical traits and three continuous traits were evaluated. Evaluations were performed at the time of physiological maturity of the coffee fruits.

The categorical traits evaluated were as follows: vegetative vigor (Vig), field evaluation of rust incidence (Rus) and cercosporiosis incidence (Cer), fruit maturation time (Mat), and fruit size (FS). The vegetative vigor was evaluated by the general appearance of the plant, by observing plant leaf development, leaf color, nutritional status, and health of coffee plants. A score scale ranging from 1 to 10 was used, where 1 was attributed to totally depleted plants and 10 was assigned to highly vigorous plants. Rust incidence and cercosporiosis were evaluated by a score scale ranging from 1 to 5, where 1 was attributed to genotypes with no symptoms of the pathogen, and 5 was assigned to highly susceptible genotypes, cercosporiosis. Fruit maturation time was classified as early, intermediate, and late, with scores ranging from 1 to 3, respectively. Fruit size was classified as small, medium, and large, with scores from 1 to 3, respectively.

The continuous traits evaluated were as follows: plant height (PH), diameter of the canopy projection (DC), and yield in liters per plant (Y). Plant height (cm) was determined by measuring the most developed orthotropic branch, from the ground to the last apical point of the coffee plant, using a measurement tape fixed to a wooden rod. The diameter of the canopy projection was determined in centimeters (cm), using a ruler perpendicular to the planting row. The yield per coffee plant was evaluated by harvesting all the fruits in a genotype and measuring the total volume in liters of freshly harvested coffee.

Analysis of phenotypic data

The phenotypes were corrected for environmental effects of years and blocks using the Selegen REML/BLUP software (de Resende 2016). The model used was as follows: $y = Xu + Za + Wc + Qs + Sb + e$, where y is the data vector; u is the vector of year-mean effects (assumed as fixed) added to the overall mean; c is the vector of specific combining ability effects between the Conilon and Robusta parents (assumed as random and distributed as $N \sim I\sigma_c^2$); a is the vector of additive genetic effects of individuals (assumed as random and distributed as $N \sim A\sigma_a^2$); s is the vector of permanent effects of individuals (assumed as random and distributed as $N \sim I\sigma_s^2$); b is the vector of permanent environment effects of blocks (assumed as random and distributed as $N \sim I\sigma_b^2$); and e is the

residual vector (assumed as random and distributed as $N \sim I\sigma_e^2$). All the effects were assumed as uncorrelated. Uppercase represent the incidence matrices for these effects. The corrected phenotypes were given by $y^* = y - Xu - Sb$ and are called deregressed phenotypes, which enter in the genomic analyses (Garrick et al. 2009; de Andrade et al. 2019).

The selective accuracy was obtained by the equation $r_{yy} = (1 - PEV/\sigma_a^2)^{1/2}$, where σ_a^2 is the additive genetic variation between individuals under evaluation and PEV is the variance of the prediction error, given by $PEV = C_i^{22}\sigma_e^2$, where C_i^{22} is the i th element of the inverse diagonal of the matrix of the coefficients of the mixed model equations, and σ_e^2 is the residual variance.

According to the model $y = Xu + Za + Wc + Qs + Sb + e$, the individual heritability was estimated by the following: $h^2 = \sigma_a^2 / (\sigma_a^2 + \sigma_c^2 + \sigma_s^2 + \sigma_b^2 + \sigma_e^2)$, where σ_j^2 is the variance component associated to the j effect.

Genomic DNA extraction, identification, and quality analysis of SNP markers

Young and fully expanded leaves of the 165 coffee trees under study were collected, and the genomic DNA was extracted using the methodology described by Diniz et al. (2005). The DNA concentration was verified in NanoDrop 2000, and its quality was evaluated in 1% agarose gel. The DNA concentration of the samples was standardized and sent to RAPiD Genomics, located in Florida, USA, for the construction of probes, sequencing, and identification of SNP molecular markers.

To identify SNP markers and coffee genotyping, 10,000 probes were selected from 40,000 polymorphic probes (Resende et al. 2016), and 18,111 SNP markers were identified. The probes were constructed from reference sequences. One of the databases was the Brazilian Genome Coffee Project, which contains over 200,000 ESTs (expressed sequence tags), corresponding to about 33,000 transcribed genes, known as Unigenes (Vieira et al. 2006). Another one was the reference genome of the *C. canephora* species, containing a total of 25,574 genes (Denoëud et al. 2014). Using these reference sequences, specific probes were obtained so that the whole genome was covered, considering both coding and non-coding regions. With these probes, the coffee genotypes were sequenced using the Illumina platform, and the SNP markers were identified using the methodology developed by the company RAPiD Genomics (Resende et al. 2016), developed for humans (Gnirke et al. 2009), and adapted to plants (Neves et al. 2013, 2014). This technology uses a method of genotyping-by-sequencing of specific regions of the genome. Details of the construction of the probes and identification of the SNP markers can be obtained in the

study carried out by Alkimim et al. (2018). The SNPs set was subject to analysis of quality implemented in the Rbio software (Bhering 2017). Quality control of SNPs was carried out by the MAF (minor allele frequency—higher than or equal to 5%) and/or call rate (CR—higher than or equal to 90%). The critical level for the MAF parameter was obtained by the equation $MAF = \frac{1}{\sqrt{2N}}$, where N refers to the total number of genotypes evaluated (de Resende et al. 2017).

Prediction using the GBLUP model

Analyses were carried out using the GBLUP method via RKHS (Reproducing Kernel Hilbert Spaces) (Gianola 2006), with a Bayesian algorithm, via R environment, in the BGLR package (Resende 2008; Perez and De Los Campos 2014). RKHS accounts for the genetic effects using the Gaussian kernel matrix (K). $K = \exp. (-hD / \text{median}(D))$, where h is the reduction coefficient to K values, h is equal to 1, and D is the Euclidean distance of codified markers matrix. A total of 100,000 Markov Chain Monte Carlo (MCMC) iterations were used, with a burn-in of the first 2000 MCMC iterations and a sampling interval (thinning) of 10.

The general mixed linear model (de Resende 2007, 2015; VanRaden 2008) was adjusted to estimate the additive genetic effects of the individuals: $y^* = Xm + Zg + e$, where y^* is the vector of corrected phenotypic observations, m is the vector of fixed effects (general mean), g is the vector of random effects of the additive genomic effects of the individuals (assumed distributed as $N \sim G\sigma_g^2$), and e refers to the vector of random residuals. Uppercase letters represent the incidence matrices for these effects. The genomic mixed model equations for the prediction of g using the GBLUP method are given by the following:

$$\begin{bmatrix} X'X & X'Z \\ Z'X & Z'Z + G\frac{\sigma_e^2}{\sigma_g^2} \end{bmatrix} \begin{bmatrix} \hat{m} \\ \hat{g} \end{bmatrix} = \begin{bmatrix} X'y^* \\ Z'y^* \end{bmatrix}$$

The genomic relationship matrix G comes from a incidence matrix M which contains the values 0, 1, and 2 for the number of alleles of the marker (or the so-called QTL) in a diploid individual.

The component M_{ij} refers to the element i of the row j of the matrix M , referring to individual j . G is a function of MM' (VanRaden 2008). The genomic heritability was computed as $h_a^2 = \frac{\sigma_g^2}{(\sigma_g^2 + \sigma_e^2)}$, where σ_g^2 is the additive genomic variance and σ_e^2 is the residual variance.

Cross-validation

The cross-validation method K -fold was used, considering $k=11$ folds. The set of observations of 165 genotypes was,

randomly, divided into groups. In the process of analysis, 150 genotypes were used as training population, and the group of 15 genotypes (remaining of original population of 165 individuals) was used as the validation population. This procedure was repeated 11 times ($k=11$) so that all groups of excluded genotypes were used in the validation.

Predictive capacity, prediction, and accuracy bias of GWS

The predictive capacity and the prediction bias are practical measures of the capacity of a method in predicting with accuracy and not with bias. The predictive capacity (r_{gy}) is determined by the correlation between the predicted genomic values and the observed phenotypic values, which are equivalent to the GWS predictive capacity to estimate the phenotypes. The prediction bias (b) is determined by the coefficient of regression of the predicted genomic values on the phenotypic values (de Resende et al. 2012; Pértile et al. 2016). The accuracy was determined by the estimator $r_{gg} = r_{gy} / \sqrt{h^2}$, where r_{gy} is the prediction ability of GWS, and h^2 is the individual heritability (Borém and Fritsche-Neto 2013).

Estimate of the number of QTL (n_{QTL}) and number of individuals (N_i) to obtain desired accuracy

The estimate of the number of QTLs that control each trait was calculated by the expression $n_{QTL} = \frac{(1-r_{gg}^2)Nh^2}{r_{gg}^2}$, where r_{gg} is equivalent to the GWS accuracy, N refers to the number of individuals in the population, and h^2 is the individual heritability (de Resende et al. 2014).

The estimate of the number of individuals (N_i) that should be evaluated in order to obtain desired accuracy was calculated by the expression $N_i = \frac{r_{gg}^2 n_{QTL}}{(1-r_{gg}^2)h^2}$, where r_{gg} is equivalent to the accuracy of GWS, n_{QTL} is the number of QTLs that control each trait, and h^2 is the individual heritability (de Resende et al. 2014).

Efficiency of GWS

The selective efficiency of GWS compared with the selection based only on 6-year phenotypes was calculated using the expression $Ef = \frac{r_{gy}L_t}{r_{yy}L_{GWS}}$, where r_{gy} is the predictive capacity of GWS, r_{yy} is the accuracy of the selection based on phenotypes, L_{jt} is the mean time required for the selection cycle based on phenotypes, and L_{GWS} is the mean time required for the selection cycle based on GWS (de Resende et al. 2012).

Table 3 Genome-wide selection (GWS) and estimates of selective accuracy based on phenotypic data obtained by mixed model analysis (REML/BLUP) for eight morphoagronomic traits in a breeding population of *Coffea canephora*

Traits	$h_a^2 \pm \text{sd}$	$r_{gy} \pm \text{sd}$	$b \pm \text{sd}$	r_{yy}	r_{gg}
Vig	0.43 ± 0.04	0.44 ± 0.15	1.18 ± 0.50	0.60	0.67
Rus	0.37 ± 0.04	0.48 ± 0.15	1.40 ± 0.67	0.39	0.79
Cer	0.43 ± 0.04	0.54 ± 0.20	1.31 ± 0.57	–	0.82
PH	0.36 ± 0.04	0.41 ± 0.22	1.35 ± 0.86	0.67	0.68
DC	0.53 ± 0.03	0.58 ± 0.13	1.23 ± 0.41	0.57	0.80
Mat	0.21 ± 0.19	−0.03 ± 0.19	1.08 ± 2.99	–	–
FS	0.21 ± 0.19	0.00 ± 0.29	0.60 ± 1.42	–	–
Y	0.15 ± 0.02	−0.02 ± 0.30	−0.11 ± 2.12	–	–

h_a^2 : genomic heritability; sd: standard deviation; r_{gy} : predictive capacity; b : predictive bias; r_{yy} : accuracy of selection based on the phenotypic data obtained by the REML/BLUP method; r_{gg} : accuracy of GWS; Vig: vegetative vigor; R: rust incidence; Cer: cercosporiosis incidence; PH: plant height; DC: diameter of the canopy projection; Mat: fruits maturation time; FS: fruit size; Y: yield per plant

Results

Analysis of phenotypic data

Phenotype data were corrected for environmental effects of years and blocks. The values of selective accuracy (r_{yy}) were estimated from the phenotypic evaluations (Table 3).

No satisfactory predictive capacity was verified for the traits fruit maturation time, fruit size, and yield per plant. Therefore, the accuracy values of these traits were not estimated. In addition, for the trait cercosporiosis incidence, selective accuracy was not estimated since the value of broad-sense heritability was 0, based on the phenotypic data.

In general, the evaluated traits showed r_{yy} of high magnitude. Values ranged from 39% for rust incidence to 67% for the plant height.

Analysis of quality of SNP markers

With the sequencing of the 165 genotypes using 10,000 probes (previously selected and distributed throughout the

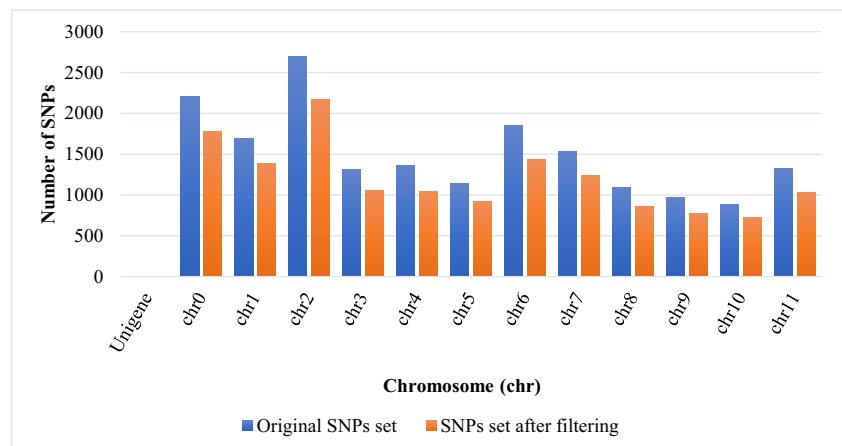
genome), 18,111 SNPs were identified. After the quality analyses, carried out in the Rbio software (Bhering 2017), 14,429 SNP markers were obtained. The initial set of SNP markers reduced by 20.33% (Fig. 1). The number of SNPs per chromosome, after the quality analyses, ranged from 4 to 2163. The highest number of SNPs was observed on chromosomes 0 and 2 (Fig. 1). We made available a file with 14,429 SNPs used in the genetic analyzes and their respective positions in the genome (Online Resource 1).

Genomic heritability, predictive capacity, prediction, and accuracy bias of GWS

Estimates of genomic heritability values, predictive capacity of GWS, prediction bias, and accuracy based on the phenotype data are shown in Table 3.

The estimated genomic heritability values (h_a^2) ranged from 0.15 for the trait yield per plant (Y) to 0.53 for the trait diameter of the canopy projection (DC). Despite the considerable h_a^2 values obtained for fruit maturation time (0.21), fruit

Fig. 1 Number of SNP markers by chromosomes obtained from original SNPs set and SNPs set after filtering. SNP markers distributed throughout the UNIGENES from the EST sequences of *Coffea arabica* and from the 11 and the chromosome 0 of *Coffea canephora*



size (0.21), and yield per plant (0.15), their predictive capacity was low.

Regarding predictive capacity of GWS (r_{gy}), the traits Vig (0.44), Rus (0.48), Cer (0.54), PH (0.41), and DC (0.58) stood out for their high estimate. This confirms that, in general, the r_{gy} values were higher for traits that had the highest h_a^2 values.

The prediction bias (b) resulted in values close to 1.0 for the traits vegetative vigor, rust incidence, cercosporiosis incidence, plant height, diameter of the canopy projection, fruit maturation time, and fruit size. The trait yield per plant showed no prediction bias close to 1.0. In addition, this trait had the lowest estimate value of genomic heritability.

The accuracy of GWS (r_{gg}) for the traits fruit maturation time, fruit size, and plant yield was not satisfactory, and therefore, their accuracy was not estimated.

The estimates of the accuracy values were obtained for the other traits, ranging from 67% (Vig) to 82% (Rus). der_{gg} values were moderate (68%) to high (79%), even for plant height and rust incidence, which had low h_a^2 values (0.36 and 0.37, respectively).

Estimate of the number of QTL (n_{QTL}) and number of individuals (Ni) to obtain desired accuracy

The estimated number of QTLs controlling each trait ranged from 35 (Cer) to 87 (Vig). In addition, the lowest values of accuracy of GWS, 67% (Vig) and 68% (PH), were obtained for the traits that had the highest number of QTLs (Table 4).

Table 4 shows the estimated number of individuals (Ni) that should be evaluated to achieve desired accuracy (r_{ggd}). The values of desired accuracy used were of 0.50, 0.60, 0.70, 0.80, and 0.90 to estimate the number of individuals. This calculation considered estimates of the genomic heritability values (h_a^2), shown in Table 3, and the estimated number of QTLs (n_{QTL}) controlling each trait, shown in Table 4. To obtain desired accuracy of 70%, which is considered of high magnitude (de Resende and Duarte 2007), 194 individuals need to be evaluated for the trait vegetative vigor, 96 for coffee rust, 78 for cercosporiosis incidence, 184 for plant height, and 89 for diameter of the canopy projection. For all traits evaluated, the higher the accuracy desired, the larger was the number of individuals to be analyzed.

Efficiency of GWS

Figure 2 shows the efficiency of the GWS with the decrease of the selective cycle in relation to the selection based only on 6 years of phenotypic data, for all traits that had good predictive capacity, except for cercosporiosis incidence (Cer). Thus, the GWS efficiency was estimated for the traits vegetative

Table 4 Number of QTLs that control the trait (n_{QTL}) and number of individuals (Ni) required to achieve desired accuracy of GWS (r_{ggd}) in a breeding population of *Coffea canephora* for Vig, Rus, Cer, PH, and DC traits

Traits	n_{QTL}	Number of individuals (Ni)				
		r_{ggd} 0.50	r_{ggd} 0.60	r_{ggd} 0.70	r_{ggd} 0.80	r_{ggd} 0.90
Vig	87	67	114	194	360	863
Rus	37	33	56	96	178	426
Cer	35	27	46	78	145	347
PH	69	68	108	184	341	817
DC	49	31	52	89	164	394

Vig: vegetative vigor; Rus: rust incidence; Cer: cercosporiosis incidence; PH: Plant height; DC: diameter of the canopy projection

vigor, rust incidence, plant height, and diameter of the canopy projection.

Figure 2 shows an increase in selective efficiency by using GWS, even for vegetative vigor and plant height, which had high estimates of accuracy from selection based on phenotypic data (60 and 67%, respectively) (Table 2). Even the traits with low h_a^2 values, plant height (0.36), and rust incidence (0.37) showed efficiency gains with GWS. With the decrease of the selective cycle from 6 to 3 years, GWS was higher (ranging from 22 to 146%) for all traits.

Discussion

Analysis of phenotypic data

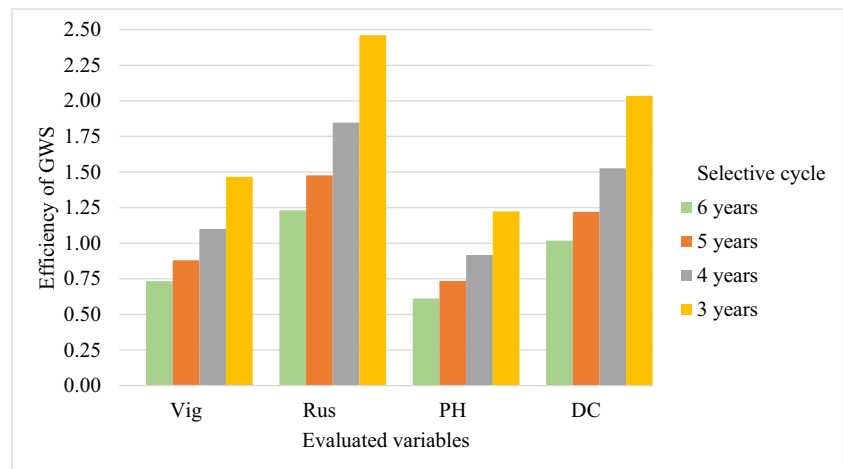
Selective accuracy (r_{yy}) was estimated by the REML/BLUP method (de Resende 2016). Selective accuracy reflects the quality of the information and procedures used to predict the genetic values of the individuals (Sousa et al. 2019).

In general, the evaluated traits had high magnitude r_{yy} , ranging from 39% (R) to 67% (PH). Therefore, the higher the value of the selective accuracy, the higher is the confidence in the evaluation and the predicted genetic value of an individual (Sousa et al. 2019).

Quality analysis of the SNP markers

Quality analyses revealed 14,429 SNP molecular markers. Quality evaluations allow identifying markers with ideal quality criteria (Sant'Ana et al. 2018). In addition, this evaluations are advantageous for they remove poor quality markers prior to the statistical analyses, consequently decreasing the occurrence of false-positive (type I error) and false-negative (type II error) (Anderson et al. 2010). High marker density is essential for capturing genes with lower and higher effect and,

Fig. 2 Efficiency of GWS in relation to selection based only on a 6-year long phenotypic data, in a breeding population of *Coffea canephora*, for the variables vegetative vigor (Vig), rust incidence (Rus), plant height (PH), and diameter of the canopy projection (DC)



consequently, increasing the probability of explaining most of the genetic variation of the study trait (Resende et al. 2008; Resende et al. 2016). Valente et al. (2016) found that the use of higher marker densities is required to obtain prediction accuracy of high magnitude (>70%, according to de Resende and Duarte 2007).

A greater number of SNPs were identified on chromosomes 0 and 2. However, chromosome 0 is not a true chromosome, but a set of unsorted sequence scaffolds of *C. canephora*. This result, number of SNPs identified on chromosome 2, may be due to the length of chromosome 2 in the genome of *C. canephora* (Denoeud et al. 2014).

Genomic heritability, predictive capacity, prediction bias, and GWS accuracy

The results of estimated values of genomic heritability (0.15 for yield per plant at 0.53 for diameter of the canopy projection) indicate the inheritable capacity of each trait. Despite the considerable h_a^2 values, given the genetic complexity of the traits, obtained for the fruit maturation time (0.21), fruit size (0.21), and yield per plant (0.15), their predictive capacity was low. Estimates of predictive capacity are expected to be lower for traits with low heritability (Legarra et al. 2008). The fact that these three traits had the lowest h_a^2 values justifies their low predictive capacity values.

Good predictive capacity (r_{gy}) was recorded for the traits Vig (0.44), Rus (0.48), Cer (0.54), PH (0.41), and DC (0.58), indicating the capacity to anticipate phenotypes for these traits. All of them are lower than 58%. Although 58% is not a high (above 70%) magnitude, it can be possible do have genetic gain from genomic selection and it can be higher per unit of time than that from phenotypic selection. These data show that, in general, the r_{gy} values were higher for the traits that had the highest h_a^2 values. A GWS study with cashew tree (*Anacardium occidentale*) also revealed a response of the

predictive capacity in function of the heritability (Cavalcanti et al. 2012).

The prediction bias (b) was close to 1.0 for vegetative vigor, rust incidence, cercosporosis, plant height, diameter of the canopy projection, fruit maturation time, and fruit size. Predictive bias close to 1.0 indicates that the prediction was non-biased and, therefore, is effective in predicting the real magnitudes of the differences between the individuals evaluated (Resende et al. 2012). The yield per plant trait had a biased prediction. This trait had also the lowest estimate value of genomic heritability, which may justify the observed bias. In addition, traits governed by larger numbers of genes require populations with larger sample sizes.

The accuracy can be classified as very high (>90%); high (70–90%); moderate (50–70%), and low (<50%) (de Resende and Duarte 2007; Rabier et al. 2016). The traits fruit maturation time, fruit size, and yield per plant showed unsatisfactory predictive capacity. Therefore, their accuracy values were not estimated. The estimates of the accuracy values were obtained for the other traits, ranging from 67% (Vig) to 82% (R). The r_{gg} values were moderate (68%) to high (79%), even for plant height and rust incidence, which had low h_a^2 values (0.36 and 0.37, respectively). These results confirm the efficiency of GWS in the selection of traits with low heritability and agree with other studies (Legarra et al. 2008; Zhang et al. 2010).

Estimate of the number of QTLs (n_{QTL}) and number of individuals (N_i) to obtain desired accuracy

The estimated number of QTLs controlling each trait ranged from 35 (Cer) to 87 (Vig). These results show the quantitative nature of the traits evaluated in this study. In addition, traits with the highest number of QTLs were those with the lowest values of accuracy of GWS, 67% (Vig) and 68% (PH). A study carried out with oil palm (*Elaeis guineensis* Jacq.) revealed that the accuracy of GWS is inversely proportional to the number of QTLs that control the traits (Wong and

Bernardo 2008). This is expected because traits governed by a larger number of QTLs are more complex. Also, in polygenic traits, in general, each gene has a little contribution to the manifestation of the trait. However, even in a scenario under low heritability and a higher number of QTLs, results of our GWS work were promising. This event probably occurred due to the set of SNPs selected by the quality criteria. Another fact is that the SNPs are widely distributed along the genome of this species. These results lead us to highlight the importance of implementing the GWS technique in the Brazilian coffee breeding program. This can be gradually incorporated, obeying the pace and conditions of each program, as an auxiliary tool in the practical conduction of improvement programs, in order to obtain quick and accurate gains.

The estimate of the number of individuals required to achieve desired accuracy was obtained considering values of desired accuracy of 0.50, 0.60, 0.70, 0.80, and 0.90. To obtain desired accuracy (r_{ggd}) of 70% (high magnitude) (de Resende and Duarte 2007), 194 individuals would have to be evaluated for vegetative vigor, 96 for coffee rust incidence, 78 for cercosporiosis incidence, 184 for plant height, and 89 for diameter of the canopy projection. Thus, this study evaluated more individuals than necessary to achieve accuracy of 70% for most of the traits (Rus, Cer, and DC). In addition, this study revealed that the higher the desired accuracy, the larger is the number of individuals to be analyzed.

Efficiency of GWS

The y -axis values shown in Fig. 2 mean the ratio of gain, per unit of time, between genomic selection, and phenotypic selection. When higher than 1, it indicates that genomic selection will provide superior gain. For example, for a value of 1.25, the gain is 25%. Genomic prediction is an additional information on the genetic value of the individual and as such can allow an earlier selection with precision and so can enable reduction of the number of harvests below four (it is well-known that coffee trials require, on average, four harvests for an efficient selection).

In perennial species, such as *C. canephora*, one of the advantages of genomic selection is the shortening of the selection cycle in order to practice early selection (Castro et al. 2016). In our work, the selective efficiency increased when using GWS, even for vegetative vigor and plant height, which showed high estimates of accuracy based only on phenotype data (60 and 67%, respectively). This increase in efficiency is due to the reduction in the time required to complete a selective cycle using GWS. Thus, reducing the cycle from 6 to 3 years increased the selective efficiency of the GWS for all the traits throughout the reduction of the cycle. Therefore, even when the accuracy of genomic selection has the same magnitude as that obtained with selection based on phenotypic

data, GWS will provide higher genetic gains due to the shorter selection cycle (Gois et al. 2016).

With the decrease of the selective cycle from 6 to 3 years, GWS was more efficient (ranging from 22 to 146%) for all traits. Thus, the reduction in the time required to complete a selective cycle is significant when using GWS. Genomic prediction and selection can be performed at the seedling stage, and therefore, GWS has a higher efficiency per unit of time (Resende et al. 2012; Gois et al. 2016). Similar results were observed in other studies. In a study evaluating the selective efficiency of GWS with a 50% decrease in the selective cycle of citrus, GWS was superior for all the traits evaluated (ranging from 31 to 160%) (Gois et al. 2016). A study with oil palm (*Elaeis guineensis* Jacq.) revealed a reduction in the selection cycle from 19 to 6 years when using GWS (Wong and Bernardo 2008). In a study with maize (*Zea mays* L.), the use of GWS significantly increased the selective accuracy and the genetic gains per unit of time (Fritsche-Neto et al. 2012). In other studies, genomic selection also showed great potential in increasing breeding efficiency by simulated results (Resende et al. 2008; Valente et al. 2016).

GWS provided gains in efficiency even for traits with low h_a^2 , plant height (0.36), and rust incidence (0.37). These data confirm the importance of GWS in the selection of low-heritability traits, a fact that is also observed in other studies (Resende et al. 2012; Gois et al. 2016).

Conclusion

Results reveal that genome-wide selection is useful for *C. canephora* breeding since it accurately predicts the phenotypes of individuals. This fact leads to a significant reduction in the time required to complete the selection cycle, providing gains in selective efficiency per unit of time. In addition, these results can be used as the basis for further studies on the genus *Coffea* and on perennial species with genetic similarity.

Data archiving statement The SNP data was included in the [Supplementary Materials](#).

Funding information This work was financially supported by the Brazilian Coffee Research and Development Consortium (Consórcio Brasileiro de Pesquisa e Desenvolvimento do Café—CBP&D/Café), by the Foundation for Research Support of the state of Minas Gerais (FAPEMIG), by the National Council of Scientific and Technological Development (CNPq), by the Coordenação de Aperfeiçoamento de Pessoal de Nível Superior—Brasil (CAPES)—Finance Code 001, and by the National Institutes of Science and Technology of Coffee (INCT/Café).

Compliance with ethical standards

Conflict of interest The authors declare that they have no conflict of interest.

References

- Alkimim ER, Caixeta ET, Sousa TV, Pereira AA, Oliveira ACB, Zambolim L, Sakiyama NS (2017) Marker-assisted selection provides arabica coffee with genes from other *Coffea* species targeting on multiple resistance to rust and coffee berry disease. *Mol Breed* 37:6–10. <https://doi.org/10.1007/s11032-016-0609-1>
- Alkimim ER, Caixeta ET, Sousa TV, Silva FL, Sakiyama NS, Zambolim L (2018) High-throughput targeted genotyping using next-generation sequencing applied in *Coffea canephora* breeding. *Euphytica* 214:50–18. <https://doi.org/10.1007/s10681-018-2126-2>
- Anderson CA, Pettersson FH, Clarke GM, Cardon LR, Morris AP, Zondervan KT (2010) Data quality control in genetic case-control association studies. *Nat Protoc* 5:1564–1573. <https://doi.org/10.1038/nprot.2010.116>
- Asins MJ, Fernandez-Ribacoba J, Bernet GP et al (2012) The position of the major QTL for *Citrus tristeza virus* resistance is conserved among *Citrus grandis*, *C. aurantium* and *Poncirus trifoliata*. *Mol Breed* 29:575–587. <https://doi.org/10.1007/s11032-011-9574-x>
- Bhering LL (2017) Rbio: a tool for biometric and statistical analysis using the R platform. *Crop Breed Appl Biotechnol* 17:187–190. <https://doi.org/10.1590/1984-70332017v17n2s29>
- Bernardo R (2010) Breeding for quantitative traits in plants. Stemma Press, Woodbury
- Borém A, Fritsche-Neto R (2013) Biotecnologia Aplicada ao Melhoramento de Plantas
- Carvalho MC d CG d, Silva DCGD (2010) Sequenciamento de DNA de nova geração e suas aplicações na genômica de plantas. *Ciência Rural* 40:735–744. <https://doi.org/10.1590/S0103-84782010000300040>
- Castro CA d O, Resende RT, Bhering LL, Cruz CD (2016) Brief history of *Eucalyptus* breeding in Brazil under perspective of biometric advances. *Ciência Rural* 46:1585–1593. <https://doi.org/10.1590/0103-8478cr20150645>
- Cavalcanti JJV, de Resende MDV, dos Santos FHC, Pinheiro CR (2012) Predição simultânea dos efeitos de marcadores moleculares e seleção genômica ampla em cajueiro. *Rev Bras Frutic* 34:840–846. <https://doi.org/10.1590/S0100-29452012000300025>
- de Almeida ÍF, Cruz CD, de Resende MDV (2016) Validação e correção de fenótipos na seleção genômica ampla. *Pesqui Agropecuária Bras* 51:1973–1982. <https://doi.org/10.1590/s0100-204x2016001200008>
- de Andrade LRB, e Sousa MB, Oliveira EJ et al (2019) Cassava yield traits predicted by genomic selection methods. *PLoS One* 14:e0224920. <https://doi.org/10.1371/journal.pone.0224920>
- de Resende MDV (2007) Matemática e Estatística na Análise de Experimentos e no Melhoramento Genético, 1st edn, Colombo
- de Resende MDV (2016) Software Selegen-REML/BLUP: a useful tool for plant breeding. *Crop Breed Appl Biotechnol* 16:330–339. <https://doi.org/10.1590/1984-70332016v16n4a49>
- de Resende MDV, Duarte JB (2007) Precisão e controle de qualidade em experimentos de avaliação de cultivares. *Pesqui Agropecuária Trop* 37:182–194
- de Resende MDV, e Silva FF, Lopes PS, Azevedo CF (2012) Seleção Genômica Ampla (GWS) via Modelos Mistos (REML/BLUP), Inferência Bayesiana (MCMC), Regressão Aleatória Multivariada e Estatística. Espacial, Viçosa
- de Resende MDV, e Silva FF, Azevedo CF (2014) Estatística Matemática, Biométrica e Computacional: Modelos Mistos, Multivariados, Categorias e Generalizados (REML/BLUP), Inferência Bayesiana, Regressão Aleatória, Seleção Genômica, QTI-GWAS, Estatística Espacial e Temporal, Competição, Sobrevivência, 1st edn. Suprema, Viçosa
- Resende MDV de, e Silva FF, Ferreira C, et al (2017) Atualidades da biometria no melhoramento de plantas perenes. In: Ludke WH, Andrade ACB, Volpato L, et al. (eds) *Desafios Biométricos No Melhoramento Genético*, 1ª edição. Viçosa, p 166
- Denoed F, Carretero-Paulet L, Dereeper A et al (2014) The coffee genome provides insight into the convergent evolution of caffeine biosynthesis. *Science* 345(80):1181–1184. <https://doi.org/10.1126/science.1255274>
- Diniz LEC, Sakiyama NS, Lashermes P et al (2005) Analysis of AFLP markers associated to the Mex-1 resistance locus in Icatu progenies. *Crop Breed Appl Biotechnol* 5:387–393. <https://doi.org/10.12702/1984-7033.v05n04a03>
- Ferrão LFV, Ferrão RG, Ferrão MAG et al (2017) A mixed model to multiple harvest-location trials applied to genomic prediction in *Coffea canephora*. *Tree Genet Genomes* 13:95–13. <https://doi.org/10.1007/s11295-017-1171-7>
- Ferrão LFV, Ferrão RG, Ferrão MAG, Fonseca A, Carbonetto P, Stephens M, Garcia AAF (2019) Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. *Heredity* (Edinb) 122:261–275. <https://doi.org/10.1038/s41437-018-0105-y>
- Fritsche-Neto R, Resende MDV, Miranda GV, DoVale JC (2012) Seleção genômica ampla e novos métodos de melhoramento do milho. *Rev Ceres* 59:794–802. <https://doi.org/10.1590/S0034-737X2012000600009>
- Garrick DJ, Taylor JF, Fernando RL (2009) Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol* 41:55. <https://doi.org/10.1186/1297-9686-41-55>
- Gianola D (2006) Genomic-assisted prediction of genetic value with semiparametric procedures. *Genetics* 173:1761–1776. <https://doi.org/10.1534/genetics.105.049510>
- Gnirke A, Melnikov A, Maguire J, Rogov P, LeProust E, Brockman W, Fennell T, Giannoukos G, Fisher S, Russ C, Gabriel S, Jaffe DB, Lander ES, Nusbaum C (2009) Solution hybrid selection with ultra-long oligonucleotides for massively parallel targeted sequencing. *Nat Biotechnol* 27:182–189. <https://doi.org/10.1038/nbt.1523>
- Goddard ME, Hayes BJ (2007) Genomic selection. *J Anim Breed Genet* 124:323–330. <https://doi.org/10.1111/j.1439-0388.2007.00702.x>
- Gois IB, Borém A, Cristofani-Yaly M et al (2016) Genome wide selection in *Citrus* breeding. *Genet Mol Res* 15:1–14. <https://doi.org/10.4238/gmr15048863>
- Grattapaglia D, Resende MDV (2011) Genomic selection in forest tree breeding. *Tree Genet Genomes* 7:241–255. <https://doi.org/10.1007/s11295-010-0328-4>
- Heslot N, Jannink J-L, Sorrells ME (2015) Perspectives for genomic selection applications and research in plants. *Crop Sci* 55:1–12. <https://doi.org/10.2135/cropsci2014.03.0249>
- Jannink J-L, Lorenz AJ, Iwata H (2010) Genomic selection in plant breeding: from theory to practice. *Brief Funct Genomics* 9:166–177. <https://doi.org/10.1093/bfgp/elq001>
- Kemper KE, Goddard ME (2012) Understanding and predicting complex traits: knowledge from cattle. *Hum Mol Genet* 21:R45–R51. <https://doi.org/10.1093/hmg/dd332>
- Lande R, Thompson R (1990) Efficiency of marker-assisted selection in the improvement of quantitative traits. *Genetics* 124:743–756
- Legarra A, Robert-Granie C, Manfredi E, Elsen J-M (2008) Performance of genomic selection in mice. *Genetics* 180:611–618. <https://doi.org/10.1534/genetics.108.088575>
- Leroy T, Marraccini P, Dufour M, Montagnon C, Lashermes P, Sabau X, Ferreira LP, Jourdan I, Pot D, Andrade AC, Glaszmann JC, Vieira LG, Piffanelli P (2005) Construction and characterization of a *Coffea canephora* BAC library to study the organization of sucrose biosynthesis genes. *Theor Appl Genet* 111:1032–1041. <https://doi.org/10.1007/s00122-005-0018-z>
- Liao P-Y, Lee KH (2010) From SNPs to functional polymorphism: the insight into biotechnology applications. *Biochem Eng J* 49:149–158. <https://doi.org/10.1016/j.bej.2009.12.021>

- Lopez GAG, McCouch SR, Moncada MDP (2013) A genetic map of an interspecific diploid pseudo testcross population of coffee. *Euphytica* 192:305–323. <https://doi.org/10.1007/s10681-013-0926-y>
- Meuwissen T (2007) Genomic selection: marker assisted selection on a genome wide scale. *J Anim Breed Genet* 124:321–322. <https://doi.org/10.1111/j.1439-0388.2007.00708.x>
- Meuwissen T, Hayes B, Goddard M (2001) Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157:1819–1829 [11290733](https://doi.org/10.1534/genetics.114.164442)
- Neves LG, Davis JM, Barbazuk WB, Kirst M (2013) Whole-exome targeted sequencing of the uncharacterized pine genome. *Plant J* 75:146–156. <https://doi.org/10.1111/tpj.12193>
- Neves LG, Davis JM, Barbazuk WB, Kirst M (2014) A high-density gene map of loblolly pine (*Pinus taeda* L.) based on exome sequence capture genotyping. *G3-Genes Genomes Genet* 4:29–37. <https://doi.org/10.1534/g3.113.008714>
- Noir S, Anthony F, Bertrand B et al (2003) Identification of a major gene (Mex-1) from *Coffea canephora* conferring resistance to *Meloidogyne exigua* in *Coffea arabica*. *Plant Pathol* 52:97–103. <https://doi.org/10.1046/j.1365-3059.2003.00795.x>
- Perez P, De Los Campos G (2014) Genome-wide regression and prediction with the BGLR statistical package. *Genetics* 198:483–495. <https://doi.org/10.1534/genetics.114.164442>
- Pértile SFN, e Silva FF, Salvian M, Mourão GB (2016) Seleção e associação genômica ampla para o melhoramento genético animal com uso do método ssGBLUP. *Pesqui Agropecuária Bras* 51:1729–1736. <https://doi.org/10.1590/s0100-204x2016001000004>
- Rabier C-E, Barre P, Asp T, Charmet G, Mangin B (2016) On the accuracy of genomic selection. *PLoS One* 11:e0156086. <https://doi.org/10.1371/journal.pone.0156086>
- Resende MDV (2008) Genômica Quantitativa e Seleção no Melhoramento de Plantas Perenes e Animais, Colombo: E
- Resende MDV, Lopes PS, Silva RL, Pires IE (2008) Seleção genômica ampla (GWS) e maximização da eficiência do melhoramento genético. *Pesqui Florest Bras* 56:63–77
- Resende MDV, Resende MFR, Sansaloni CP et al (2012) 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. <https://doi.org/10.1111/j.1469-8137.2011.04038.x>
- Resende M, Caixeta E, Alkimim ER et al (2016) High-throughput targeted genotyping of *Coffea arabica* and *Coffea canephora* using next generation sequencing, San Diego, p 1
- Resende MDV, Ramalho MAP, Guilherme SR, de F.B. Abreu Â (2015) Multigeneration Index in the Within-Progenies Bulk Method for Breeding of Self-pollinated Plants. *Crop Sci* 55:1202–1211. <https://doi.org/10.2135/cropsci2014.08.0580.z>
- Romero G, Vásquez LM, Lashermes P, Herrera JC (2014) Identification of a major QTL for adult plant resistance to coffee leaf rust (*Hemileia vastatrix*) in the natural Timor hybrid (*Coffea arabica* x *C. canephora*). *Plant Breed* 133:121–129. <https://doi.org/10.1111/pbr.12127>
- Sant'Ana GC, Pereira LFP, Pot D et al (2018) Genome-wide association study reveals candidate genes influencing lipids and diterpenes contents in *Coffea arabica* L. *Sci Rep* 8:465. <https://doi.org/10.1038/s41598-017-18800-1>
- Sousa TV, Caixeta ET, Alkimim ER, Oliveira ACB, Pereira AA, Zambolim L, Sakiyama NS (2017) Molecular markers useful to discriminate *Coffea arabica* cultivars with high genetic similarity. *Euphytica* 213:1–15. <https://doi.org/10.1007/s10681-017-1865-9>
- Sousa TV, Caixeta ET, Alkimim ER et al (2019) Early selection enabled by the implementation of genomic selection in *Coffea arabica* breeding. *Front Plant Sci*:9. <https://doi.org/10.3389/fpls.2018.01934>
- Toppa EVB, Jadoski CJ (2013) O uso dos marcadores moleculares no melhoramento genético de plantas. *Sci Agrar Parana* 12:1–5. <https://doi.org/10.18188/1983-1471>
- Valente MSF, Viana JMS, de Resende MDV et al (2016) Seleção genômica para melhoramento vegetal com diferentes estruturas populacionais. *Pesqui Agropecuária Bras* 51:1857–1867. <https://doi.org/10.1590/s0100-204x2016001100008>
- Van Eenennaam AL, Weigel KA, Young AE et al (2014) Applied animal genomics: results from the field. *Annu Rev Anim Biosci* 2:105–139. <https://doi.org/10.1146/annurev-animal-022513-114119>
- VanRaden PM (2008) Efficient methods to compute genomic predictions. *J Dairy Sci* 91:4414–4423. <https://doi.org/10.3168/jds.2007-0980>
- Vieira LGE, Andrade AC, Colombo CA et al (2006) Brazilian coffee genome project: an EST-based genomic resource. *Brazilian J Plant Physiol* 18:95–108. <https://doi.org/10.1590/S1677-04202006000100008>
- 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. <https://doi.org/10.1007/s00122-008-0715-5>
- Zambolim L (2016) Current status and management of coffee leaf rust in Brazil. *Trop Plant Pathol* 41:1–8. <https://doi.org/10.1007/s40858-016-0065-9>
- Zhang Z, Liu J, Ding X et al (2010) Best linear unbiased prediction of genomic breeding values using a trait-specific marker-derived relationship matrix. *PLoS One* 5:e12648. <https://doi.org/10.1371/journal.pone.0012648>
- Zhao Y, Mette MF, Reif JC (2015) Genomic selection in hybrid breeding. *Plant Breed* 134:1–10. <https://doi.org/10.1111/pbr.12231>

Publisher's note Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.