

Genome selection in fruit breeding: application to table grapes

Alexandre Pio Viana^{1*}, Marcos Deon Vilela de Resende², Summaira Riaz³, Michael Andrew Walker³

¹State University of North Fluminense "Darcy Ribeiro" – Genetic and Plant Breeding Lab., Av. Alberto Lamego, 2000 – Campos dos Goytacazes, RJ – Brazil.

²Embrapa Forestry, Estrada da Ribeira, Km 111 – 83411-000 – Colombo, PR – Brazil.

³University of California – Dept. of Viticulture & Enology, One Shields Av. – 95616-5270 – Davis, CA – USA.

*Correspondent author <pirapora@uenf.br>

Edited by: Martin Boer

Received September 18, 2014

Accepted September 01, 2015

ABSTRACT: Genomic selection (GS) has recently been proposed as a new selection strategy which represents an innovative paradigm in crop improvement, now widely adopted in animal breeding. Genomic selection relies on phenotyping and high-density genotyping of a sufficiently large and representative sample of the target breeding population, so that the majority of loci that regulate a quantitative trait are in linkage disequilibrium with one or more molecular markers and can thus be captured by selection. In this study we address genomic selection in a practical fruit breeding context applying it to a breeding population of table grape obtained from a cross between the hybrid genotype D8909-15 (*Vitis rupestris* × *Vitis arizonica/girdiana*), which is resistant to dagger nematode and Pierce's disease (PD), and 'B90-116', a susceptible *Vitis vinifera* cultivar with desirable fruit characteristics. Our aim was to enhance the knowledge on the genomic variation of agronomical traits in table grape populations for future use in marker-assisted selection (MAS) and GS, by discovering a set of molecular markers associated with genomic regions involved in this variation. A number of Quantitative Trait Loci (QTL) were discovered but this method is inaccurate and the genetic architecture of the studied population was better captured by the BLasso method of genomic selection, which allowed for efficient inference about the genetic contribution of the various marker loci. The technology of genomic selection afforded greater efficiency than QTL analysis and can be very useful in speeding up the selection procedures for agronomic traits in table grapes.

Keywords: MAS, SSR markers, linkage disequilibrium, genomic selection

Introduction

Over the last 10 years, the genome sequence of about 20 plant species including *Vitis* have been made publicly available. There are more than 850 simple sequence repeat (SSR) markers in the grape data base, and the grape genome sequence provides insight into the evolution of this globally important fruit crop, and is now being used to speed up the development of new varieties. Availability of genome sequence information along with high throughput genotyping platforms is changing the nature of research experiments in the effort to understand the evolution of organisms, as well transform the strategies for genetic improvement. One artificial selection strategy based on genome-wide markers, called genomic selection, is revolutionizing the genetic improvement of animals and plants species (Kumar et al., 2012a).

Genomic selection (GS) has recently been proposed as a new selection strategy to compare to MAS based on a few identified QTL. It is an innovative paradigm in crop improvement, having become widely adopted in animal breeding. Genomic selection relies on phenotyping and high-density genotyping of a sufficiently large and representative sample of the target breeding population, so that the majority of loci that regulate a quantitative trait are in linkage disequilibrium with one or more molecular markers (Resende et al., 2012). In contrast to MAS, in genomic selection, the effects of all available genetic markers are estimated simultaneously in a training population, and models are developed to predict the genomic value of breeding of progeny in future generations (Meuwissen et al., 2001).

In this study we address genomic selection in a practical fruit breeding context by applying it to a breeding population of table grape obtained by a cross between the hybrid genotype D8909-15 (*Vitis rupestris* × *Vitis arizonica/girdiana*), which is resistant to dagger nematode and Pierce's disease (PD), and 'B90-116', a susceptible *Vitis vinifera* cultivar with desirable fruit characteristics. Our aim was to enhance the knowledge on the genomic variation of agronomical traits in table grape populations for future use in marker-assisted selection (MAS) and GS, by discovering a set of molecular markers associated with genomic regions involved in this variation.

Materials and Methods

Mapping, training and selection population

An F₁ population called "AT0023" of 203 individuals was developed from a cross between D8909-15 (female) and *V. vinifera* B90-116 (male, a seedless table grape cultivar) that were highly resistant and susceptible to PD, respectively. D8909-15 was derived from a cross *V. rupestris* A. de Serres × *V. arizonica* b42-26. The maternal grant parent line was also highly susceptible to PD, while the paternal grant parent was strongly resistant to PD. Out of the 203 progeny, 143 were evaluated for agronomical traits and were used for construction of genomic maps, QTL analysis and marker effect estimation adopting the GS approach.

Phenotype traits

Five clusters were randomly chosen from each genotype and following Viana et al., (2013) a qualitative

and quantitative evaluation of the total number of clusters (NC) per genotype was made. A leaf score (LS) was determined by visual comparison to the attributes of each parent, and then scores from 1 to 5 were recorded. The D8909-15 parent, which appeared more *V. rupestris*-like, was given a score of 1 (based on the depth and width of the petiolar sinus) and the B90-116 *V. vinifera* parent was given a score of 5 (also based on the shape of the petiolar sinus). Scoring of the progeny was based on where they fell within this continuum. The length of peduncle (LP) was obtained by measuring the peduncle on each of the collected clusters. The length of cluster to peduncle (CL) was obtained by measuring the collected clusters from the tip to the peduncle insertion. The number of berries (NB) was obtained from each of the collected clusters. The weight of 10 berries (WB) was obtained by weighing on an electronic scale. The soluble solids (Brix per cluster) were determined by a portable refractometer and recorded as °Brix. Juice samples were also prepared from 25 berries collected from 3 different clusters, by squashing the berries and filtering the juice through nylon mesh. The following measurements were taken: juice pH was determined using a calibrated pH meter (Corning pH meter 430), titratable acidity (TA) was determined in a 5 mL juice sample to which three drops of 1 % phenolphthalein indicator were added, and then titrated, under agitation, with 0.1 N NaOH solution standardized beforehand with potassium biphthalate.

Genome maps and QTL analyses

Genomic DNA was extracted from 143 genotypes of the AT0023 population along with parental lines D8909-15, B90116, A. de Serres and b42-26 according to a published protocol (Riaz et al., 2004). We used 243 SSR markers, and the source of SSR markers is the same as described previously. PCR amplification of SSR markers and visual scoring were also similarly conducted as explained earlier (Riaz et al., 2006).

The double pseudo-testcross strategy (Grattapaglia and Sederoff, 1994) and JoinMap 3.0 software (Plant Research International, Wageningen, Netherlands) were used to build genetic maps. Markers with high distortion or unexpected chi-square test results were discarded. Linkage groups were determined using the Kosambi function for the translation of recombination units into genetic distance. An LOD score threshold for determination of linkage groups was 3.5. The recombination fraction permitted was 0.45. Markers within the resulting groups were ordered relative to each other by automatic multipoint analyses using the default values of JoinMap 3.0 (mapping threshold $LOD > 1$, recombination frequency threshold < 0.4).

A consensus map was constructed using the parameters for a cross-pollinated derived population and the integrate map function of JoinMap 3.0. The linkage groups were numbered according to the reference map of Riaz et al. (2004) and the international agreement

achieved within the IGGP (International Grape Genome Program; www.vitaceae.org).

We applied the Hidden Markov models to the QTL analysis using multiple imputation by interval mapping (Broman and Sen, 2009) as described by Viana et al., 2013, where each genotype is imputed randomly, but conditional on the observed marker genotype data: we simulated from the joint genotype distribution given the observed data.

We estimated the μ_j and σ by maximum likelihood; that is, we took as our estimates those values for which the observed data was most probable. The likelihood function was:

$$L(\mu_j, \sigma) = \prod_i \sum_j P_{ij} \Phi(y_i, \mu_j, \sigma^2)$$

where the sum is above the possible QTL genotypes. We used a form of the expectation-maximization (EM) algorithm beginning with initial estimates $\hat{\mu}_j^0$ and $\hat{\sigma}^{2(0)}$.

Once the maximum likelihood estimates of the μ_j and σ had been obtained, a LOD score was calculated as follows:

$$LOD = \log_{10} \left(\frac{\prod_i \sum_j P_{ij} \Phi(y_i; \hat{\mu}_j, \hat{\sigma}^2)}{\prod_i \Phi(y_i; \hat{\mu}_0, \hat{\sigma}_0^2)} \right)$$

where $\hat{\mu}_0$ and $\hat{\sigma}_0$ were the average and standard deviation (SD) of the y_i , so that the denominator of the LOD score is the likelihood under the null hypothesis that there is no QTL anywhere in the genome. A LOD score threshold for determination of QTL presence was 4.2.

GS approach

RR-BLUP method

The RR-BLUP/GS method (Meuwissen et al., 2001) uses predictors of the BLUP type, but the effects of markers are set as explanatory variables rather than classificatory variables. Thus, they are regressive variables adjusted as covariables of random effects, i.e., the phenotypes are regressed based on these covariables.

The estimators associated with the random regression or ridge regression promote shrinkage controlled by a function of the penalization parameter λ . When λ is not known, its arbitrary choice leads to the "ridge regression" (RR) method. If the regression parameter is associated to $\lambda = \sigma_e^2 / \sigma_{gi}^2 = \sigma_e^2 / (\sigma_g^2 / n_Q)$, there is a BLUP random regression for the effect of chromosomal segment i , where σ_{gi}^2 is the additive genetic variance associated with the locus or segment i ; and σ_g^2 and σ_e^2 are the additive genetic variance of the character and the residual variance, respectively. The penalty parameter λ can also be determined by iterative means or fine tuning, by choosing the one that maximizes the correlation between the phenotypic value and genetic estimates breeding value (GEBV) predicted in a cross-validation population (Meuwissen et al., 2001).

The distinction between fixed regression, ridge regression and random regression in a model using only phenotypes is associated with the penalization parameter λ^* , which is given by $\lambda = (1-h^2)/h^2$. Small λ^* values are enough to reduce the impact of this multicollinearity present among the covariables in the $Z'Z$ matrix, which is approximately singular. λ^* value equal to zero (h^2 value equal to 1) characterizes the fixed regression. Small λ^* values (0.01 to 0.1) characterize the ridge regression and high λ^* values (higher than 0.1) characterize the random regression (Resende et al., 2012).

Prediction with RR-BLUP is described below based on the following general linear mixed model that is set for estimating the effects of markers: $y = Xb + Zm + e$, where y is the vector of phenotypic observations; b is the vector of fixed effects; m is the vector of the random effects of markers; and e refers to the vector of random residuals. X and Z are the incidence matrices for b and m , respectively. For SSR markers, the incidence matrix Z contains the values 0 and 1, which are derived from the absence/presence of one allele of the marker (or alleged QTL) in a diploid individual. This is similar to what is used with dominant markers such as DarT (Pérez et al., 2010). The genomic mixed model equations for predicting m by the RR-BLUP method is equivalent to (Resende et al., 2012):

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

The genomic estimated breeding value (GEBV) of an individual j belonging to the validation population is predicted as a sum of the k locus contribution by:

$$\hat{y}_j = \sum_k Z_{jk} \hat{m}_k$$

This genomic estimated breeding value (GEBV) is equivalent to the total genomic genetic value of the individual j .

The prediction equations presented above assume *a priori* that all gene loci explain equal amounts of genetic variation. Therefore, the genetic variation explained by each gene locus is given by σ_g^2/n_Q , where σ_g^2 is the total genetic variation and n_Q is the number of genes (when each locus is perfectly marked by a single marker) controlling the trait and can be given by

$$n_Q = \sum_{i=1}^n p_i q_i$$

where p_i is the allelic frequency of M in locus and n is the number of marker loci (Habier et al., 2007). The genetic variation σ_g^2 can be estimated by REML traditionally, according to the phenotypic data, or by variation itself between the markers or QTL chromosome segments.

BLASSO

In the Bayesian Lasso (Park and Casella, 2008; Campos et al., 2009) the prior assigned to marker effects m_a (with incidence matrix W) is a Laplace (double exponential, DE) distribution. All marker effects are assumed to be independently and identically distributed as DE. This prior assigns the same variance or prior uncertainty to a collection of marker effects, but it possesses thicker tails than the normal or Gaussian prior and as a result the variance σ_{mai}^2 of an additive marker effect is specific to each marker locus i .

With the scalar u referring to the intercept and residual effects e , the model is in the following form:

$$\begin{aligned} y &= 1u + Zm_a + e \\ e | \sigma_e^2 &\sim MVN(0, I\sigma_e^2) \\ m_a | \lambda, \sigma_e^2 &\sim \prod_i (\lambda / 2\sigma_e) \exp[(-\lambda |m_{ai}|) / \sigma_e] \end{aligned}$$

Using a formulation in terms of an augmented hierarchical model, we have:

$$\begin{aligned} p(m_{ai} | \tau) &\sim N(0, D\sigma_e^e); \text{diag}(D) = (\tau_1^2 \dots \tau_n^2) \\ p(\tau | \lambda) &= \prod_i (\lambda^2 / 2) \exp(-\lambda^2 \tau_i^2 / 2). \end{aligned}$$

Then the genetic variance in one locus is $\sigma_{mai}^2 = \tau_i^2 \sigma_e^2$ and the residual variance is σ_e^2 .

The form of the distribution of marker effects is determined by the penalization parameter λ , which is related to the marker genetic variation by the expression $\text{Var}(m_a) = 2\sigma_e^2 / \lambda^2$. This relation denotes that λ^2 plays a similar role as the inverse of the variance in the Gaussian models. BLASSO was fitted through a BLR package in R (de los Campos et al., 2009; Pérez et al., 2010).

Cross validation

The generalized Jackknife method was used and was based on dividing the set of N sample points into g groups of size k , such that $N = gk$. In general, k is set at 1. The estimates were based on samples of size $(g-1)k$, where the i -th group of size k was removed. For $k=1$, $N=g$ and $(g-1)k = g-1 = N-1$, such that the estimate m_i is obtained in the sample in which the observation i was omitted. The average over all groups furnishes the final estimate of the marker effect. The procedure was applied to different sets of markers elected according to their retained number (5, 10, 50, 100, 200, 243) defined by the greatest moduli of their estimated marker effects (Usai et al., 2009).

The number of markers was chosen with the aim of giving a reasonable curve of the behavior of predictive ability as a function of marker number. The pattern of results was consistent and not erratic, showing no ups and downs in the subsequent number of markers. This confirms that the intermediate number of markers does not yield higher predictive ability than the extremes of the interval, except at its maximum.

Estimates of the genomic selection's efficiency of predictive ability

To calculate the efficiency of GS, genomic values of individuals of the validation population are predicted (using the estimated effects from the estimation population) and subjected to correlation analysis with their observed phenotypic values. As the validation sample was not involved in predicting the marker effects, the errors arising from genomic values and phenotypic values are independent. Thus, the correlations between these values are predominantly genetic in nature and equivalent to the predictive ability (r_{gp}) of GS in estimating phenotypes, which is related to the accuracy of selection itself. Coefficients of determination of the genomic breeding value estimation were given by r_{gp}^2 (Usai et al., 2009).

Regression of Phenotypes on Predictions

The regression coefficient involving observed and predicted values is a practical measure of the ability of the methods to make predictions that are unbiased.

The regression coefficient is algebraically equal to 1. Regression coefficients less than 1 indicate that the genetic values are overestimated and exhibit greater variability than expected; coefficients greater than 1 indicate that the estimated genetic values exhibit less than expected variability. Lack of bias is important when selection involves individuals from many generations using estimated marker effects from a single generation. Regression coefficients near 1 indicate that the assessments are unbiased and effectively predict the actual magnitudes of differences between the individuals assessed (Resende et al., 2012).

Efficiency of GS over QTL analyses

The coefficient of determination of the QTL effects (R^2_{QTL}) was given as the proportion of phenotypic variation and is explained by the QTL. The efficiency of GS was measured as the number of times that GS is more efficient than MAS, based on the QTL mapped and was estimated as the relation between the determination coefficients provided by the respective methods (Efficiency = R^2_{QTL} / R^2_{pg}). Both of these quantities, R^2_{QTL} and R^2_{pg} , are proportions explained by the phenotypic variation. As such, they are explaining the same total variation and so are comparable and can be used to make inferences about efficiency.

Results

QTL description

Quantitative trait loci, identified by Viana et al., 2013 using interval mapping and Hidden Markov models are presented in Table 1. The number of QTL detected for each trait varied between one and eight, reflecting the quantitative nature of these traits. The total percentage of variation explained by these QTL was small to medium, varying between 4 % and 25 %. QTL were observed in different linkage groups (LG). Eight QTL were

Table 1 – Characteristics of the detected QTLs (Some Quantitative Trait Locus) for each trait measured in segregant population of table grapes derived from a (*Vitis rupestris* × *V. arizonica/girdiana*) × *V. vinifera* population (Viana et al., 2013).

| Trait | Number of linkage group | Number of markers – QTL's | % of variation |
|-------------------------------|-------------------------|---------------------------|----------------|
| Number of cluster | 08 | 08 | 25 |
| Leaf score | 03 | 03 | 8 |
| Brix per cluster | 03 | 03 | 9 |
| Tartaric acid | 03 | 03 | 5 |
| Length of peduncle | 03 | 03 | 11 |
| Length of cluster to peduncle | 01 | 01 | 4 |
| Number of berries | 03 | 03 | 8 |
| Wt. of 10 berries | 04 | 04 | 16 |

identified for number of cluster across LG1, 5, 6, 7, 12, 13, 14 and 19. Three QTL were detected for leaf score across LG3, 5 and 14, for brix per cluster we found QTL in LG1, 6 and 11. For tartaric acid we found QTL in three different LG.

Three QTL were detected regarding length of peduncle across LG9, 10 and 12 and number of berries (LG4, 9 and 14), and one QTL was observed for length of cluster to peduncle (LG14). For the weight of 10 berries we found four QTL (LG1, 8, 10 and 11).

GS approach

The RR-BLUP and BLasso genomic selection methods were compared (Table 2) using 243 SSR markers. For the RR-BLUP, regression of phenotypes on predictions (Beta values) were all below the desirable ones (close to 1). For all variables analyzed, optimal (around 1) Beta values were obtained for 50 markers added to the model using the BLasso method (Tables 2 and 3). This number (50) of markers also maximized the predictive ability with the BLasso method. Thus, this method was superior to the RR-BLUP method.

The RR-BLUP method for the trait number of cluster with 50 SSR markers, gave a beta value of 0.66; the beta values for leaf score, brix per cluster, tartaric acid, length of peduncle, length of cluster to peduncle, number of berries and weight of 10 berries were 0.77, 0.65, 0.77, 0.65, 0.65, 0.66 and 0.71, respectively. The correlation coefficients between the genomic breeding values and phenotypic values ranged between 0.26 and 0.72 (Table 2). The values that best reflect this correlation were obtained for genomic sampling with 50 markers. For number of cluster, a correlation of 0.66 was estimated for the other variables (Leaf score, brix per cluster, tartaric acid, length of peduncle, length of cluster to peduncle, number of berries and weight of 10 berries) and the correlations were 0.77, 0.65, 0.77, 0.65, 0.65, 0.66 and 0.71, respectively. Values close to unity (1.0) are ideal for this correlation.

For the BLasso methodology, the best Beta values were also estimated with 50 SSR markers. These values ranged between 0.89 and 0.98, which demonstrates the

Table 2 – Number of markers retained in the model, average predictive ability (correlation r_{pg}) and regression of phenotypes on predictions (beta) by RR-BLUP and Bayesian Lasso (Blasso) methods for agronomical traits in table grapes.

| Trait | Nmarkers | Beta-RR | r_{pg} -RR | Beta-Blasso | r_{pg} -Blasso |
|-------------------------------|----------|---------|--------------|-------------|------------------|
| Number of cluster | 5 | 0.76 | 0.38 | 0.83 | 0.38 |
| | 10 | 0.79 | 0.51 | 0.87 | 0.51 |
| | 50 | 0.66 | 0.58 | 0.94 | 0.65 |
| | 100 | 0.54 | 0.45 | 0.99 | 0.58 |
| | 200 | 0.58 | 0.21 | 1.19 | 0.20 |
| | 243 | - | -0.05 | - | -0.16 |
| Leaf score | 5 | 0.88 | 0.54 | 0.91 | 0.54 |
| | 10 | 0.85 | 0.57 | 0.92 | 0.57 |
| | 50 | 0.77 | 0.67 | 0.98 | 0.72 |
| | 100 | 0.68 | 0.54 | 1.04 | 0.64 |
| | 200 | 0.60 | 0.29 | 1.34 | 0.37 |
| | 243 | 0.42 | 0.13 | 1.05 | 0.14 |
| Brix per cluster | 5 | 0.82 | 0.45 | 0.87 | 0.45 |
| | 10 | 0.84 | 0.56 | 0.91 | 0.56 |
| | 50 | 0.65 | 0.55 | 0.94 | 0.62 |
| | 100 | 0.59 | 0.47 | 0.97 | 0.56 |
| | 200 | 0.44 | 0.17 | 1.00 | 0.20 |
| | 243 | - | -0.02 | - | -0.03 |
| Tartaric acid | 5 | 0.74 | 0.36 | 0.85 | 0.35 |
| | 10 | 0.95 | 0.61 | 0.91 | 0.62 |
| | 50 | 0.77 | 0.72 | 0.93 | 0.77 |
| | 100 | 0.63 | 0.62 | 0.96 | 0.75 |
| | 200 | 0.59 | 0.46 | 0.94 | 0.62 |
| | 243 | 0.51 | 0.28 | 1.19 | 0.42 |
| Length of peduncle | 5 | 0.80 | 0.42 | 0.86 | 0.42 |
| | 10 | 0.78 | 0.49 | 0.89 | 0.49 |
| | 50 | 0.65 | 0.51 | 0.97 | 0.57 |
| | 100 | 0.67 | 0.48 | 1.11 | 0.55 |
| | 200 | 0.48 | 0.15 | 0.95 | 0.15 |
| | 243 | - | -0.06 | - | -0.12 |
| Length of cluster to peduncle | 5 | 0.84 | 0.46 | 0.87 | 0.46 |
| | 10 | 0.81 | 0.54 | 0.88 | 0.54 |
| | 50 | 0.65 | 0.56 | 0.89 | 0.62 |
| | 100 | 0.46 | 0.36 | 0.95 | 0.51 |
| | 200 | 0.36 | 0.13 | 1.02 | 0.17 |
| | 243 | - | -0.09 | - | -0.15 |
| Number of berries | 5 | 0.88 | 0.55 | 0.91 | 0.55 |
| | 10 | 0.95 | 0.61 | 0.92 | 0.61 |
| | 50 | 0.66 | 0.60 | 0.91 | 0.68 |
| | 100 | 0.66 | 0.48 | 0.94 | 0.62 |
| | 200 | 0.44 | 0.26 | 1.06 | 0.38 |
| | 243 | 0.26 | 0.08 | 0.89 | 0.12 |
| Wt. of 10 berries | 5 | 0.84 | 0.47 | 0.88 | 0.47 |
| | 10 | 0.81 | 0.55 | 0.89 | 0.55 |
| | 50 | 0.71 | 0.63 | 0.95 | 0.69 |
| | 100 | 0.67 | 0.57 | 0.98 | 0.67 |
| | 200 | 0.52 | 0.21 | 1.17 | 0.25 |
| | 243 | - | -0.14 | - | -0.23 |

potential of this methodology for estimating genomic breeding values for this population structure in table grapes. Values above 1.00 were estimated for various traits when 100 markers were added to the estimation models, which revealed bias in the estimates. The correlation coefficients between the phenotypic and genomic breeding values were also higher for this method, namely, 0.65, 0.72, 0.62, 0.77, 0.57, 0.62, 0.68, 0.69 for the following variables: number of berries, leaf score, brix per cluster, tartaric acid, length of peduncle, length of cluster to peduncle, number of berries and the weight of 10 berries, respectively. Considering the estimation of genomic breeding values by the Blasso methodology, Table 3 shows the values already mentioned and the comparison with the analyses of QTL. The determination coefficients of the genomic breeding values (proportion of phenotypic variance explained by the 50 SSR markers) are higher than the determination coefficients of the estimated QTLs' effects, for the following variables: number of cluster (0.42 and 0.25 respectively), leaf score (0.52 and 0.08), brix per cluster (0.38 and 0.09), tartaric acid, length of peduncle, length of cluster to peduncle, number of berries and the weight of 10 berries, 0.59 and 0.05, 0.32 and 0.11, 0.38 and 0.05, 0.46 and 0.08, 0.48 and 0.16, respectively. When the two techniques are compared, estimated relative efficiencies are much higher and range from 1.7 to 11.6. In other words, genomic selection efficiency in the variables analyzed was between 1.7 and 11.6 times higher than the approach by QTL analysis.

Discussion

MAS by QTL analysis

In a study developed by Viana et al., 2013, several QTL affecting agronomical traits in table grapes were identified. Clusters of QTL for closely related traits in several linkage groups were observed. QTL for the number of fruit cluster and the weight of 10 berries were found in several linkage groups (1, 4, 6, 7, 10, 11, 12, 13, 14, 19). For these two traits there is significant correlation, and one marker explains 8 % of the total phenotypic variation (VVMD25). In another marker there is different and low variation for the number of clusters, in this case, reflecting the quantitative nature for this trait. These results differ in part from QTL previously published by Fanizza et al., (2005) and Doligez et al., (2010), QTL for similar traits in different linkage groups were found. This discrepancy could result from a segregation difference between crosses (Viana et al., 2013). Part of this divergence might also be attributed to differences in trait measurement.

QTL for length of peduncle and length of cluster to peduncle in different linkage groups (9, 10, 12, 14), in this study had no correlation between them (Viana et al., 2013). For the trait length of cluster to peduncle significant correlation with a number of berries (0.664**) were found, and in the same linkage group QTL for the

Table 3 – Optimal predictive ability (R_{pg}) according to the best prediction method (BLasso), coefficient of determination of the genomic model (R^2_{pg}) and of the QTLs (Some Quantitative Trait Locus) model r^2_{QTLs} as well as the regression of phenotypes on predictions (Beta).

| Trait | Best method | Number of markers | Beta | r_{pg} | r^2_{pg} | r^2_{QTLs} | Relative efficiency |
|-------------------------------|-------------|-------------------|------|----------|------------|--------------|---------------------|
| Number of cluster | BLasso | 50 | 0.94 | 0.65 | 0.42 | 0.25 | 1.7 |
| Leaf score | BLasso | 50 | 0.98 | 0.72 | 0.52 | 0.08 | 6.5 |
| Brix per cluster | BLasso | 50 | 0.94 | 0.62 | 0.38 | 0.09 | 4.1 |
| Tartaric acid | BLasso | 50 | 0.93 | 0.77 | 0.59 | 0.05 | 11.6 |
| Length of peduncle | BLasso | 50 | 0.97 | 0.57 | 0.32 | 0.11 | 3.0 |
| Length of cluster to peduncle | BLasso | 50 | 0.89 | 0.62 | 0.38 | 0.05 | 7.8 |
| Number of berries | BLasso | 50 | 0.91 | 0.68 | 0.46 | 0.08 | 6.0 |
| Wt. of 10 berries | BLasso | 50 | 0.95 | 0.69 | 0.48 | 0.16 | 3.0 |

number of berries were found (9, 14). If all these factors are taken into consideration, it is logical to imagine that a large number of genes and different physiological mechanisms might be involved in the determination of each fruit trait in response to yearly environmental variations.

For the weight of 10 berries we found QTL across linkage groups 1, 8, 10 and 11. Other investigations have detected stable QTL for berry weight in table grapes and wine grapes; however, the QTL detected in these investigations were all different. These different results might be due to the different progenies used (Doligez et al., 2002; Fischer et al., 2004; Fanizza et al., 2005). This variability in berry weight in the different progenies might have affected the detection of the QTL and together with progeny size and heritability, might play an important part in explaining the different results. Small sample size and medium-to-low trait heritability might have biased QTL detection (Beavis, 1998; Melchinger et al., 2004).

For the fruit quality of fruit, several QTL were found, which explains moderate values for the phenotypic variation. For brix per cluster QTL across linkage groups 1 and 3 were found, and in linkage group 3 the SSR marker (VMC1a5), explains 9 % for the total phenotypic variation. For this trait we found negative correlation with Titratable acidity (-0.647**) was found. For the Titratable acidity we found QTL across linkage groups 6, 13 and 19 and they explained low variation.

There were no reports of QTL analysis for traits related to the fruit quality of grape (Brix and Titratable acidity). The location of QTL in different linkage groups suggests that the genetic control of these characters are influenced by several genes involved in complex metabolic pathways, so a higher saturation of the linkage groups obtained should be performed for the localization of QTL with major effect.

These results may highlight the difficulties in the implementation of marker-assisted selection by QTL detection and mapping for agronomic traits. The low variation explained by few markers directly affects selection models, which leads to low efficiency in the use of the technique. For perennial crops, it leads to low capacity for application and implementation in practical fruit breeding programs aiming to accelerate the procedures for selecting superior genotypes.

Implementing genomic selection in table grape breeding programs

A few studies reported on the use of breeding values predicted through RR-BLUP and BLasso methods in fruit crops with the aim of developing a genomic selection approach. Such studies have shown that using these breeding values in association with the phenotype resulted in higher accuracy when selecting new superior genotypes, supporting the approach used in our study (Kumar et al., 2012b; Iwata et al., 2013).

By means of genomic selection, seedlings can be genotyped and selected for different agronomical traits, initiating a new breeding cycle. Under this approach, rapid pyramiding of favorable alleles can be pursued by establishing crosses that create the best allelic complement across quantitative trait loci throughout the genome. In parallel, selected seedlings can be clonally replicated and established in clonal trials to verify their performance relative to elite commercial genotypes. Because a breakdown in accuracy in genomic selection is known to occur across generations, the application of this methodology will require monitoring of the accuracy of prediction models and their recalibration after a few generations, as necessary (Resende et al., 2012).

The prediction models proposed by various authors are now preferred in studies on the association between markers and phenotypic values, since they introduced the concept of simultaneous prediction of the effects of markers, without using significance tests for individual markers. This approach differs from the procedures primarily recommended for marker assisted selection (MAS). Thus, there is an emphasis on the use of the concept of linkage disequilibrium at the population level and not just within families.

GS is applied over all loci simultaneously and is based on estimation and prediction rather than hypothesis testing. This way, it can explain much more about genetic variability and prevent the so-called missing or lost heritability, typical of studies on linkage and association. This possibility is of great significance for quantitative traits in table grapes. Variables such as number of cluster, which showed high value for genomic predictive ability (r_{gp}), can be very useful in the selection of new high yield genotypes. For the trait

brix per cluster, high values of r_{sp} were also observed. This is important when aiming to increase sugar levels, which is a relevant trait for table grapes. Such high correlation coefficients between genomic breeding values and phenotypic values characterize a favorable condition for the application of this technique in segregating populations of table grapes.

This analysis shows that the models adjusted by BLasso can estimate better marker effects and explain more accurately phenotypes and are thus preferred for use in marker-assisted breeding in table grapes. In the original Lasso method, a joint mode is estimated and most markers are expected to have exactly the same effect as zero (Usai et al., 2009). In Bayesian Lasso (BLasso), averages are estimated *a posteriori* and provide many very small marker effects, but different from zero. In other words, BLasso does not physically remove markers but effectively does it. Besides, these *a posteriori* averages are excellent selection criteria. In the original Lasso method, the solution admits up to (N-1) non-zero regression coefficients, where N is the number of individuals. BLasso neglects this constraint, possibly producing a more accurate estimated model (Campos et al., 2009).

In RR-BLUP methods, prediction equations assume *a priori* that all loci explain equal amounts of genetic variation. This might not match with the genetic architecture of the traits in apple. This may partly explain the low adjustment of the explanatory models of the variables measured in this population.

Compared to the RR-BLUP, the *a priori* density used in the BLasso method presented distributions with greater mass point at values equal to zero and stronger tails, thus experiencing greater shrinkage in the regression coefficients near zero and less shrinkage in the regression coefficients far from zero (Resende et al., 2013).

The genetic architecture of the studied population was well captured by the BLasso method and allowed for efficient inference about the genetic contribution of the various marker loci. The technology of genomic selection provided greater efficiency than QTL analysis and can be very useful in speeding up the selection procedures for agronomic traits in table grapes.

Acknowledgments

The first author is grateful to Brazilian National Council for Scientific and Technological Development (CNPq) for a fellowship that allowed for training in the Department of Viticulture and Enology, University of California at Davis.

References

- Beavis, W.D. 1998. QTL analyses: power, precision and accuracy. p. 145-162. In: Paterson, A.H., ed. Molecular dissection of complex traits. CRC Press, Boca Raton, FL, USA.
- Broman, K.W.; Sen, S. 2009. A Guide to QTL Mapping with R/QTL. Springer, New York, NY, USA.
- de los Campos, G.; Naya, H.; Gianola, D.; Crossa, J.; Legarra, A. 2009. Prediction quantitative traits with regression models for dense molecular markers and pedigree. *Genetics* 182: 375-385.
- Doligez, A.; Bouquet, A.; Danglot, Y.; Lahogue, F.; Raiz, S.; Meredith, C.P.; Edwards, K.J.; This, P. 2002. Genetic mapping of grapevine (*Vitis vinifera* L.) applied to the detection of QTLs for seedlessness and berry weight. *Theoretical and Applied Genetics* 105: 780-795.
- Doligez, A.; Bertrand, Y.; Dias, S.; Grolier, M.; Ballester, J.F.; Bouquet, A.; This, P. 2010. QTLs for fertility in table grapes (*Vitis vinifera* L.). *Tree Genetics & Genomes* 6: 413-422.
- Fanizza, G.; Lamaj, F.; Costantini, L.; Chabane, R.; Grando, M.S. 2005. QTL analysis for fruit yield components in table grapes (*Vitis vinifera*). *Theoretical and Applied Genetics* 111: 658-664.
- Fischer, B.M.; Salakhutdinov, I.; Akkurt, M.; Eibach, R.; Edwards, K.J.; Topfer, R.; Zyprian, E.M. 2004. Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. *Theoretical and Applied Genetics* 108: 501-515.
- Grattapaglia, D.; Sederoff, R. 1994. Genetic linkage maps of *Eucalyptus grandis* and *Eucalyptus urophylla* using a pseudotestcross: Mapping strategy and RAPD markers. *Genetics* 137: 1121-1137.
- Habier, D.; Fernando, R.L.; Dekkers, J.C.M. 2007. The impact of genetic relationship on genome-assisted breeding values. *Genetics* 117: 2389-2397.
- Iwata, H.; Hayashi, T.; Terakami, S.; Takada, N.; Saito, T.; Yamamoto, T. 2013. Genomic prediction of trait segregation in a progeny population: a case study of Japanese pear (*Pyrus pyrifolia*). *BMC Genetics* 14: 81.
- Kumar, S.; Chagné, D.; Bink, M.C.A.M.; Volz, R.K.; Whitwork, C. 2012a. Genomic selection for fruit quality traits in apple (*Malus × domestica* Borkh). *PLoS ONE* 7: e36674. doi:10.1371/journal.pone.0036674.
- Kumar, S.; Bink, M.C.A.; Volz, R.K.; Vincent, G.M.; Chagné, D. 2012b. Towards genomic selection in apple (*Malus × domestica* Borkh) breeding programmes: prospects, challenges and strategies. *Tree Genetics & Genomes* 8: 1-14. doi: 10.1007/s11295-011-0425-z.
- Melchinger, A.E.; Utz, H.F.; Schon, C.C. 2004. QTL analyses of complex traits with cross validation, bootstrap and other biometrical methods. *Euphytica* 137: 1-11.
- Meuwissen, T.H.E.; Hayes, B.J.; Goddard, M.E. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics* 157: 1819-1829.
- Park, T.; Casella, G. 2008. The bayesian LASSO. *Journal of the American Statistical Association* 103: 681-686.
- Pérez, P.; Campos, G.; Crossa, J.; Gianola, D. 2010. Genomic-enabled prediction based on molecular markers and pedigree using the bayesian linear regression package in R. *The Plant Genome* 3: 106-116.
- Riaz, S.; Dangi, G.S.; Edwards, K.J.; Meredith, C.P. 2004. A microsatellite marker based framework linkage map of *Vitis vinifera* L. *Theoretical and Applied Genetics* 108: 864-872.
- Riaz, S.; Krivanek, A.F.; Xu, K.; Walker, M.A. 2006. Refined mapping of the Pierce's disease resistance locus, PdR1, and Sex on an extended genetic map of *Vitis rupestris* × *V. arizonica*. *Theoretical and Applied Genetics* 113: 1317-1329.

- Resende, R.M.S.; Casler, M.D.; Resende, M.D.V. 2013. Genomic selection in forage breeding: accuracy and methods. *Crop Science* 54: 143-147.
- Resende, M.D.V.; Resende, M.F.R.; Sansaloni, C.P.; Petroli, C.D.; Missiaggia, A.A.; Aguiar, A.M.; Abad, J.M.; Takahashi, E.K.; Rosado, A.M.; Faria, D.A.; Pappas, G.J.; Kilian, A.; Gratapaglia, D. 2012a. Genomic selection for growth and wood quality in Eucalyptus: capturing the missing heritability and accelerating breeding for complex traits in forest trees. *New Phytologist* 194: 116-128.
- Resende, M.F.R.; Muñoz, P.; Acosta, J.J.; Peter, G.F.; Davis, J.M.; Grattapaglia, D.; Resende, M.D.V.; Kirst, M. 2012b. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* 193: 617-624.
- Resende, M.F.R.; Muñoz, P.; Resende, M.D.V.; Garrick, D.J.; Fernando, R.L.; Davis, J.M.; Jokela, E.J.; Martin, T.A.; Peter, G.F.; Kirst, M. 2012c. Accuracy of genomic selection methods in a standard data set of Loblolly Pine (*Pinus taeda* L.). *Genetics* 190: 1503-1510.
- Usai, M.G.; Goddard, M.E.; Hayes, B.J. 2009. LASSO whit cross-validation for genomic selection. *Genetic Research* 91: 427-436.
- Viana, A.P.; Riaz, S.; Walker, M.A. 2013. Genetic dissection of agronomic traits within a segregating population of breeding table grapes. *Genetics and Molecular Research* 12: 951-964.