

Harnessing multivariate, penalized regression methods for genomic prediction and QTL detection of drought-related traits in grapevine

Charlotte Brault ^{1,2,3}, Agnès Doligez^{2,3,*}, Le Cunff^{1,2,3}, Aude Coupel-Ledru ⁴, Thierry Simonneau⁴, Julien Chiquet⁵, Patrice This^{2,3} and Timothée Flutre^{6,*}

¹Institut Français de la Vigne et du Vin, Montpellier F-34398, France,

²UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier F-34398, France,

³UMT Geno-Vigne®, IFV-INRAE-Institut Agro, Montpellier F-34398, France,

⁴LEPSE, Univ Montpellier, INRAE, Institut Agro, Montpellier 34000, France,

⁵AgroParisTech, UMR MIA, Paris 75005, France, and

⁶Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE—Le Moulon, Gif-sur-Yvette 91190, France

*Corresponding authors: Université Paris-Saclay, INRAE, CNRS, AgroParisTech, GQE—Le Moulon, Gif-sur-Yvette 91190, France. Email: timothee.flutre@inrae.fr (T.F.); UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro, Montpellier F-34398, France. Email: agnes.doligez@inrae.fr (A.D.)

Abstract

Viticulture has to cope with climate change and to decrease pesticide inputs, while maintaining yield and wine quality. Breeding is a key lever to meet this challenge, and genomic prediction a promising tool to accelerate breeding programs. Multivariate methods are potentially more accurate than univariate ones. Moreover, some prediction methods also provide marker selection, thus allowing quantitative trait loci (QTLs) detection and the identification of positional candidate genes. To study both genomic prediction and QTL detection for drought-related traits in grapevine, we applied several methods, interval mapping (IM) as well as univariate and multivariate penalized regression, in a bi-parental progeny. With a dense genetic map, we simulated two traits under four QTL configurations. The penalized regression method Elastic Net (EN) for genomic prediction, and controlling the marginal False Discovery Rate on EN selected markers to prioritize the QTLs. Indeed, penalized methods were more powerful than IM for QTL detection across various genetic architectures. Multivariate prediction did not perform better than its univariate counterpart, despite strong genetic correlation between traits. Using 14 traits measured in semi-controlled conditions under different watering conditions, penalized regression methods proved very efficient for intra-population prediction whatever the genetic architecture of the trait, with predictive abilities reaching 0.68. Compared to a previous study on the same traits, these methods applied on a denser map found new QTLs controlling traits linked to drought tolerance and provided relevant candidate genes. Overall, these findings provide a strong evidence base for implementing genomic prediction in grapevine breeding.

Keywords: genomic prediction; QTL detection; multi-trait; breeding; candidate gene; water stress; grapevine

Introduction

Viticulture is facing two major challenges, i.e., coping with climate change and decreasing inputs such as pesticides, while maintaining high yield and quality. This requires understanding the physiological processes that determine adaptation to climate change, such as water use efficiency and their genetic basis (Condon et al. 2004). Breeding schemes could then use crosses between genotypes with high water use efficiency, and others resistant to downy and powdery mildews (Vezzulli et al. 2019b), to select offspring displaying the most favorable combinations. In crop species, the widespread use of molecular markers through marker-assisted selection (MAS) or genomic prediction (GP) substantially accelerates genetic gains as compared to the traditional phenotypic selection, by allowing early selection of promising genotypes, without the corresponding phenotypic information (Heffner et al. 2009). This is of acute interest in perennial fruit

species because of the long juvenile period during which most traits of interest cannot be phenotyped. MAS and GP are now widely developed in many perennial species such as pear (Kumar et al. 2019), oil palm (Cros et al. 2015; Kwong et al. 2017), citrus (Gois et al. 2016), apple (Muranty et al. 2015), and *Coffea* (Ferrão et al. 2019). In grapevine, quantitative trait loci (QTL) detection in bi-parental populations led to the identification of major genes for traits with a simple genetic architecture such as resistance to downy and powdery mildews, berry color, seedlessness, and Muscat flavor (Fischer et al. 2004; Welter et al. 2007; Fournier-Level et al. 2009; Emanuelli et al. 2010; Mejía et al. 2011; Schwander et al. 2012). Based on these results, most breeding efforts in grapevine use MAS to improve disease resistance. However, genetic improvement is also needed for traits with a more complex genetic determinism, as well as for others, such as drought-related traits, that are difficult to phenotype. Many

Received: January 11, 2021. Accepted: July 02, 2021

© The Author(s) 2021. Published by Oxford University Press on behalf of Genetics Society of America.

This is an Open Access article distributed under the terms of the Creative Commons Attribution License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted reuse, distribution, and reproduction in any medium, provided the original work is properly cited.

minor QTLs have been found for tolerance to abiotic stresses (Marguerit *et al.* 2012; Coupel-Ledru *et al.* 2014, 2016), yield components (Doligez *et al.* 2010, 2013), and fruit quality (Huang *et al.* 2012), as reviewed in Vezzulli *et al.* (2019a). But MAS is not well suited to traits with many underlying minor QTLs (Bernardo 2008). Genomic prediction, which relies on high-density genotyping, is a promising tool for breeding for such complex traits, especially in perennial plants (Kumar *et al.* 2012). Nevertheless, in grapevine, GP has been used in three published papers, only once on experimental data (Viana *et al.* 2016a; Migicovsky *et al.* 2017) and once on simulated data (Fodor *et al.* 2014). Thus, before applying GP to this species, it has to be empirically validated by thoroughly investigating the efficiency of different methods on traits with various genetic architectures.

Both QTL detection and genomic prediction rely on finding statistical associations between genotypic and phenotypic variation. So far, QTL detection in grapevine has been mainly achieved by using interval mapping (IM) methods in bi-parental populations, or more recently genome-wide association studies (GWAS) in diversity panels [see Vezzulli *et al.* (2019a) for a comprehensive review of QTL detection studies in grapevine]. However, most quantitative traits are explained by many minor QTLs, which are difficult to detect either by IM methods or GWAS where each QTL has to exceed a significance threshold. In contrast, GP methods, by focusing on prediction, are less restrictive on the number of useful markers, sometimes resulting in all markers being retained as predictive with a nonzero effect. This is why GP methods are more efficient at predicting genotypic values (Goddard and Hayes 2007) and therefore increasingly popular with breeders (Heffner *et al.* 2010; Crossa *et al.* 2017; Kumar *et al.* 2020).

Widely used methods for GP are based on penalized regression (Hastie *et al.* 2009), notably RR [Ridge Regression, equivalent to Genomic BLUP, GBLUP, Habier *et al.* (2007)] and LASSO (Least Absolute Shrinkage and Selection Operator). Bayesian approaches are also commonly used (*e.g.*, de los Campos *et al.* 2013; Kemper *et al.* 2018), see Desta and Ortiz (2014) for a classification of GP methods. However, overall, Bayesian methods do not achieve better predictive ability than RR or LASSO, while they bear a heavy computational cost when fitted using Markov chain Monte-Carlo algorithms (Ferrão *et al.* 2019). Other methods based on nonparametric models (*e.g.*, Support Vector Machine, Reproducing Kernel Hilbert Space, Random Forest) have been shown to yield lower predictive ability than parametric models (frequentist or Bayesian) when the trait displayed an additive genetic architecture (Azodi *et al.* 2019).

Traits are often analyzed one by one in GP, using univariate methods. Nevertheless, breeders want to select the best genotypes that combine good performance for many favorable traits. Analyzing several traits jointly in GP allows taking into account any genetic correlation between traits (Henderson and Quaas 1976). Calus and Veerkamp (2011), Jia and Jannink (2012), Hayashi and Iwata (2013), and Guo *et al.* (2014) compared univariate vs multivariate models' performance. They found a slight advantage for multivariate analysis when heritability was low and data were missing. Predictive ability was particularly improved if the test set had been phenotyped for one trait while prediction was applied to another correlated trait (trait-assisted prediction) as in Thompson and Meyer (1986), Jia and Jannink (2012), Pszczola *et al.* (2013), Lado *et al.* (2018), Velazco *et al.* (2019), and Liu *et al.* (2020). However, this breaks independence between the training and test sets, leading to over-optimistic prediction accuracy (Runcie and Cheng 2019). Multivariate methods have also been proposed for QTL detection by Jiang and Zeng (1995), Korol

et al. (1995), Meuwissen and Goddard (2004), notably for distinguishing between linkage and pleiotropy when a QTL is found common to several traits. Some methods of multivariate penalized regression, such as in Chiquet *et al.* (2017), were designed to allow both QTL detection and genotypic value prediction. Multivariate GP methods are expected to perform better if traits are genetically correlated, but this remains to be confirmed with additional data. We also hypothesize that these methods will have higher QTL detection power, by making better use of information on the genetic architecture of several intertwined traits.

Methods designed for QTL detection are rarely used for genotypic value prediction. As they select only the largest QTLs, we hypothesize that these methods will provide an accurate prediction so long as the genetic architecture is simple, but would result in poor prediction performance otherwise, as determined in several studies (Heffner *et al.* 2011; Wang *et al.* 2014; Arruda *et al.* 2016). Conversely, some methods for GP, such as the LASSO and its extensions, are also able to select markers with nonnull effects, and hence to perform QTL detection. Their accuracy in detecting QTLs has been partially investigated in an inbred species by Li and Sillanpää (2012) on a single trait and simulated data and by Cho *et al.* (2010) on human data and a binary trait. Additional analyses are thus clearly needed.

This article aims to compare the ability of various methods to predict genotypic values and to detect QTLs in a bi-parental grapevine progeny, by focusing on traits related to climate change adaptation. We first complemented the only available, low density, SSR genetic map (Huang *et al.* 2012) by restriction-assisted DNA sequencing, to construct a saturated SNP map. Using this map, we then simulated phenotypic data to compare several univariate and multivariate methods and assess the impact of simulation parameters. Finally, we reanalyzed the phenotypic data on water stress from Coupel-Ledru *et al.* (2014, 2016), obtained in semi-controlled conditions. The same genotyping data and methods as those applied to simulated data were compared, providing deeper insight into the genetic determinism of key traits underlying water use efficiency, by finding new QTLs and candidate genes.

Materials and methods

Plant material

This study was based on a pseudo-F1 progeny of 188 offspring from a reciprocal cross made in 1995 between *Vitis vinifera* L. cultivars Syrah and Grenache (Adam-Blondon *et al.* 2005).

GBS markers

Genotyping was done by sequencing was performed after genomic reduction, using RAD-sequencing technology, with *ApeKI* restriction enzyme (Elshire *et al.* 2011), as described in Flutre *et al.* (2020). Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies. This yielded a total number of 17,298 SNPs.

Consensus genetic map

The genetic map was built with Lep-MAP3 (Rastas 2017), as described in <https://doi.org/10.15454/QEDX2V>. The resulting map had 3961 fully-informative markers (abxcd segregation) without missing data (missing marker genotypes being automatically imputed in Lep-MAP3). These data were numerically recoded in biallelic doses (0,1,2) according to the initial biallelic segregation and phase (Supplementary Table S1).

Simulation

Phenotype simulations were carried out to (i) compare several methods for prediction accuracy, and (ii) assess the efficacy of these methods to select the markers most strongly associated with trait variation.

Two traits, \mathbf{y}_1 and \mathbf{y}_2 were jointly simulated according to the following bivariate linear regression model: $\mathbf{Y} = \mathbf{X}\mathbf{B} + \mathbf{E}$, where \mathbf{Y} is the $n \times k$ matrix of traits, \mathbf{X} the $n \times p$ design matrix of allelic effects, \mathbf{B} the $p \times k$ matrix of allelic effects, and \mathbf{E} the $n \times k$ matrix of errors. For \mathbf{X} , the 3961 SNP markers mapped for the SxG progeny were encoded in four additive and four dominance effects. Therefore $n = 188$, $k = 2$, and $p = 31,688$. For \mathbf{B} , allelic effects corresponding to s additive QTLs were drawn from a matrix-variate Normal distribution, $\mathbf{B} \sim MV(0, \mathbf{I}, \mathbf{V}_B)$, with \mathbf{I} the $p \times p$ identity matrix and \mathbf{V}_B the $k \times k$ genetic variance-covariance matrix between traits such that $\mathbf{V}_B = \begin{bmatrix} \sigma_{B_1}^2 & \rho_B \sigma_{B_1} \sigma_{B_2} \\ \rho_B \sigma_{B_1} \sigma_{B_2} & \sigma_{B_2}^2 \end{bmatrix}$, where ρ_B is the genetic correlation among traits and $\sigma_{B_1}^2$ and $\sigma_{B_2}^2$ are the genetic variances for both traits \mathbf{y}_1 and \mathbf{y}_2 . In the same way, $\mathbf{E} \sim MV(0, \mathbf{I}, \mathbf{V}_E)$, with the $k \times k$ error variance-covariance matrix $\mathbf{V}_E = \begin{bmatrix} \sigma_{E_1}^2 & \rho_E \sigma_{E_1} \sigma_{E_2} \\ \rho_E \sigma_{E_1} \sigma_{E_2} & \sigma_{E_2}^2 \end{bmatrix}$, where ρ_E is the residual error correlation between traits, and $\sigma_{E_1}^2$ the error variance. We set ρ_B to 0.8, $\sigma_{B_1}^2$ and $\sigma_{B_2}^2$ to 0.1, ρ_E to 0, and narrow-sense heritability to 0.1, 0.2, 0.4 or 0.8, and $\sigma_{E_1}^2$ was deduced.

To explore different genetic architectures, we simulated $s = 2$ or 50 additive QTLs, located at s SNP markers, so that all corresponding additive allelic effects had nonzero values in \mathbf{B} . Because all allelic effects were drawn from the same distribution, all QTLs had “major” or “minor” effects for $s = 2$ and $s = 50$, respectively. All dominant allelic effects were set to zero. Two QTL distributions across traits were also simulated. For the first one, called “same,” all QTLs were at the same markers for both traits. For the second one, called “diff,” the two traits had no QTL in common. Thus, there was genetic correlation among traits only for the “same” QTL distribution.

For each configuration (2 or 50 QTLs, combined with “same” or “diff” distribution), the simulation procedure was replicated $t = 10$ times, each with a different seed for the pseudo-random number generator, resulting in different QTL positions and effects.

In a first simulation set, narrow-sense heritability was assumed equal for both traits and prediction was done with all methods described below. In a second set, we simulated two traits with different heritability values (0.1 and 0.5), for the “same” QTL distribution with $s = 20$ and $s = 200$ QTLs, with QTL effects drawn from a matrix-variate distribution with $\sigma_B^2 = 1$ and $\rho_B = 0.5$, in order to test the simulation parameters from Jia and Jannink (2012) with our genotyping data. For this second simulation set, prediction was done with a subset of methods only. Simulation parameters are summarized in Table 1.

Experimental design, phenotyping, and statistical analysis

Seven phenotypes related to drought tolerance had already been measured for 2 consecutive years on the Syrah \times Grenache progeny (on 186 genotypes among the existing 188), in semi-controlled conditions on the PhenoArch platform (https://www6.montpellier.inrae.fr/lepse_eng/M3P, last accessed on 07-21-21) in

Table 1 Parameter values in two sets of simulation of two traits in a bi-parental population

Simulation parameter	Same heritability values	Different heritability values
QTL number	2–50	20–200
Heritability value	0.8/0.8–0.4/0.4–0.2/0.2–0.1/0.1	0.1/0.5
Genetic variance	0.1/0.1	1/1
Genetic correlation	0.8	0.5
QTL distribution	Same-Diff	Same

Montpellier, France, as detailed in Coupel-Ledru et al. (2014, 2016). Briefly, of all replicates (six and five per genotype respectively in 2012 and 2013), three (in 2012), or two (in 2013) were maintained under well-watered conditions (well-watered condition, WW), whereas the other three were submitted to a moderate water deficit (water deficit condition, WD). Specific transpiration, i.e. transpiration rate per leaf area unit, was measured during daytime (TrS) and night-time (TrS_{night}). Midday leaf water potential (ψ_M , PsiM) was also measured and the difference between soil and leaf water potential ($\Delta\psi$, DeltaPsi) calculated. Soil-to-leaf hydraulic conductance on a leaf area basis (KS) was calculated as the ratio between TrS and DeltaPsi. Growth rate (DeltaBiomass) was estimated by image analysis. Transpiration efficiency (TE) was calculated over a period of 10 to 15 days as the ratio between growth and total water loss through transpiration during this period.

These seven phenotypes were studied under each watering condition (WW and WD). We thus considered 14 traits in this study, a trait being defined as a phenotype \times watering condition combination, and used the raw data available online (<https://doi.org/10.15454/YTRKV6>). For each trait, a linear mixed model was fitted with R/lme4 version 1.1-21 (Bates et al. 2014) using data from both years:

$$y = \mu + Y + R + x_g + y_g + x_c + y_c + O + C + D + G + G:Y + G:D + \epsilon \quad (1)$$

First, model 1 with nine fixed effects (Y year, R replicate, x_g , y_g coordinates in the platform within the greenhouse, x_c , y_c coordinates in the controlled-environment chamber where PsiM and TrS were measured, O operator for PsiM measurements, C controlled-environment chamber and D date of measurement) and three random effects (G genotype, $G : Y$ genotype-year, and $G : D$ genotype-date interactions) was fitted with maximum likelihood (ML). The R/lme4 output was given to R/lmerTest version 3.1-2 (Kuznetsova et al. 2017) to use its function “step.” Backward elimination of random-effect terms was performed using likelihood ratio test, followed by backward elimination of fixed-effect terms using F -test for all marginal terms, i.e., terms that can be dropped from the model while respecting the hierarchy of terms in the model (Kuznetsova et al. 2017).

The final model after backward elimination was then fitted with restricted maximum likelihood (ReML) to obtain unbiased estimates of the variance components and empirical BLUPs (Best Linear Unbiased Predictions) of the genotypic values. The acceptability of underlying assumptions (homoscedasticity, normality, independence) was visually assessed by plotting residuals and BLUPs. Broad-sense heritability on a genotype-mean basis was computed assuming a balanced design [see the introduction of Piepho and Möhring (2007)], as:

$$H^2 = \frac{\sigma_G^2}{\sigma_G^2 + \frac{\sigma_{G:Y}^2}{\bar{n}_{year}} + \frac{\sigma_e^2}{\bar{n}_{year} \times \bar{n}_{rep}}}, \quad (2)$$

with $\sigma_{G:Y}^2$ the genotype-year interaction variance, σ_e^2 the residual variance, \bar{n}_{year} the arithmetic mean number of trials (years) and \bar{n}_{rep} the mean number of replicates per trial. Its coefficient of variation was estimated by bootstrapping with R/lme4 and R/boot packages.

Comparison of genotypic BLUPs

We first recomputed genotypic BLUPs from the raw phenotypic data of Coupel-Ledru et al. (2014, 2016) in order to control the model selection step in a reproducible way. These new BLUPs had a strong linear correlation (>0.9) with those used in Coupel-Ledru et al. (2014, 2016), as shown in Supplementary Figure S2. The estimates of broad-sense heritability followed the same trend as in Coupel-Ledru et al. (2014, 2016) (Supplementary Figure S3). They were higher in WD condition than in WW condition for all traits except DeltaBiomass.

Genetic correlation between traits varied widely, some absolute correlation values being very high (e.g., up to 0.99 between PsiM and DeltaPsi in both conditions) when traits derived from others (Supplementary Figure S4).

Interval mapping methods

Two univariate IM methods were compared, using R/qrtl version 1.46-2 (Broman et al. 2003). For both, the probability of each genotypic class was first inferred at markers and every 0.1 cM between markers along with the genetic map, using the R/qrtl::calcgenoprob function.

Simple interval mapping:

Simple interval mapping (SIM, Lander and Botstein 1989) assumes that there is at most one QTL per chromosome. A LOD score was computed every 0.1 cM with R/qrtl::scanone, then 1000 permutations were performed to determine the LOD threshold so that the family-wise (genome wide) error rate (FWER) was controlled at 5%.

Multiple interval mapping:

Multiple interval mapping (MIM, Kao et al. 1999) allows the simultaneous detection of several QTLs. It was performed with R/qrtl::stepwiseqtl, using a forward/backward selection of Haley-Knott regression model (Haley and Knott 1992), with a maximum number of QTLs set to 4 (or 10 for ROC curve construction, see below), replicated 10 times to overcome occasional instability issues. Only main effects were included (no pairwise QTL × QTL interaction). The LOD threshold was computed with permutations (1000 for QTL detection and 10 for cross-validation of GP, see below) to determine the main penalty with R/qrtl::scantwo. QTL positions and effects were determined with R/qrtl::refineqtl and R/qrtl::fitqtl, respectively. For both methods, QTL positions were determined as those of LOD peaks above the threshold, with LOD-1 confidence intervals (Lander and Botstein 1989).

Penalized regression methods:

Genomic prediction can be seen as a high-dimension regression problem with more allelic effects (in **B**) to estimate than observations (in **Y**), known as the “ $n \ll p$ ” problem. The likelihood of such models must be regularized and various extensions, called penalized regression of the Ordinary Least Squares (OLS) algorithm

were proposed. Such penalization generally induces a bias in the estimation of allelic effects.

Univariate methods

Ridge regression:

Ridge regression (RR, Hoerl and Kennard 1970) adds to the OLS a penalty on the effects using the L2 norm and solves the following equation: $\hat{\beta}_{RR} = \text{argmin}_{\lambda} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_2^2$. As a result, all estimated allelic effects are shrunk toward zero, yet none is exactly zero. The amount of shrinkage is controlled by a regularization parameter (λ). We tuned it by cross-validation using the cv.glmnet function of the R/glmnet package version 3.0-2 (Friedman et al. 2010) with default parameters, except family = “gaussian” and $\alpha = 0$, keeping the λ value that minimizes the mean square error (MSE). Note that effects associated with correlated predictors are averaged so that they are close to identical, for a high level of regularization.

Least absolute shrinkage and selection operator:

Least absolute shrinkage and selection operator (LASSO, Tibshirani 1996) adds to the OLS a penalty on the effects using the L1 norm and solves the following equation: $\hat{\beta}_{LASSO} = \text{argmin}_{\lambda} \|Y - X\beta\|_2^2 + \lambda \|\beta\|_1$. As a result, some allelic effects are exactly equal to zero, while others are shrunk toward zero. Hence LASSO performs predictor selection, i.e., provides a sparse solution of predictors included in the best model, in addition to estimating their allelic effect. The LASSO regularization parameter (λ) was tuned by cross-validation with cv.glmnet function (family = “gaussian,” $\alpha = 1$), as above. If $n < p$, LASSO selects at most n predictors.

Extreme gradient boosting:

We first applied LASSO for dimension reduction and then Extreme Gradient Boosting, a popular machine learning method (Mason et al. 1999), to estimate marker effects. Hence, we called this method LASSO.GB. As gradient boosting is a nonlinear method, it can take into account any nonlinear interaction between markers, providing better prediction. Briefly, Extreme Gradient Boosting iteratively updates the estimation of weak predictors, in order to reduce the loss. This method requires an optimization of many parameters associated with the loss function (MSE). This optimization was done with train function from R/caret package version 6.0-86 (Kuhn 2008) using the “xgbTree” method. As the optimization of numerous parameters was computationally heavy, we fixed some of them (nrounds = max_depth = 2, colsample_bytree = 0.7, gamma = 0, min_child_weight = 1 and subsample = 0.5), while testing a grid of varying parameters (nrounds = 25, 50, 100, 150; eta = 0.07, 0.1, 0.2).

Elastic net:

Elastic net (EN, Zou and Hastie 2005) adds to the OLS both L1 and L2 penalties, the balance between them being controlled by a parameter (α); it solves the following equation: $\hat{\beta}_{EN} = \text{argmin}_{\lambda} \|Y - X\beta\|_2^2 + (1 - \alpha)\lambda \|\beta\|_2^2 + \alpha\lambda \|\beta\|_1$. Both α and λ were tuned by nested cross-validation: 20 values of α were tested between 0 and 1 and, for each of them, we used cv.glmnet function to choose between 500 values of λ . EN performs predictor selection but is less sparse than LASSO.

Note that RR, LASSO, and EN all assume a common variance for all allelic effects.

Multivariate methods

Multi-task group-EN:

Multi-task group-EN (MTV_EN, [Hastie and Qian 2016](#)) is a multivariate extension of EN, it solves the following equation: $B_{MTV_{EN}} = \operatorname{argmin}_{\lambda} \|Y - XB\|_F^2 + (1 - \alpha)\lambda\|\beta\|_F^2 + \alpha\lambda\|\beta\|_2$, F being the Frobenius norm. It assumes that each predictor variable has either a zero or nonzero effect across all traits, allowing nonzero effects to have different values among traits. λ and α parameters were tuned using `cv.glmnet` (family = "mgaussian"). MTV_RR is the multivariate extension of RR, also tuned with `cv.glmnet` (family = "mgaussian," $\alpha = 0$). MTV_LASSO is the multivariate extension of LASSO, also tuned with `cv.glmnet` (family = "mgaussian," $\alpha = 1$). The implementation of these three methods is identical.

The multivariate structured penalized regression:

The multivariate structured penalized regression (called SPRING in [Chiquet et al. 2017](#)) applies a $L1$ - penalty (λ_1 parameter) for controlling sparsity (like LASSO) and a smooth $L2$ - penalty (λ_2 parameter) for controlling the amount of structure among predictor variables (L) to add in the model, i.e., the correlation between markers according to their position on the genetic map. Both parameters λ_1 and λ_2 were tuned by cross-validation using `cv.spring` function (from `R/spring` package, version 0.1-0). The regression equation can be written as: $Y = XB + \epsilon$ with $\epsilon \sim \mathcal{N}(0, R)$, R is the covariance matrix of residuals (Gaussian noise). The allelic effects are: $B = -\Omega_{xy}\Omega_{yy}^{-1}$ and they comprise both direct effects Ω_{xy} and indirect ones Ω_{yy} .

SPRING solves the following equation: $(\hat{\Omega}_{xy}, \hat{\Omega}_{yy}) = \operatorname{argmin} -\frac{1}{n} \log \ell(\hat{\Omega}_{xy}, \hat{\Omega}_{yy}) + \frac{\lambda_2}{2} \operatorname{tr}(\hat{\Omega}_{yx}L\hat{\Omega}_{xy}\hat{\Omega}_{yy}^{-1}) + \lambda_1\|\hat{\Omega}_{xy}\|_1$. Unlike multi-task group-EN, SPRING selects specific predictors for each trait, i.e., a selected predictor can have a nonzero effect for a subset of the traits. Moreover, SPRING allows the distinction between direct and indirect effects by using conditional Gaussian graphical modeling. These effects are due to covariance of the noise such as environmental effects affecting several traits simultaneously. This distinction results in two kinds of estimated allelic effects: the direct ones, re-estimated with OLS, which are best suited for QTL detection (we called the corresponding prediction method `spring.dir.ols`) and the regression ones, which involve both direct and indirect effects and are best suited for prediction (`spring.reg` method).

Robust extension for marker selection

To enhance the reliability of marker selection by penalized methods, we used two approaches: stability selection (SS) ([Meinshausen and Bühlmann 2009](#)) and marginal False Discovery Rate ([Brehehy 2019](#)), both of which aim at controlling the number of false-positive QTLs. We did not use these methods for genomic prediction, as they are not designed for this purpose.

Stability selection:

SS is a method that controls FWER, it computes the empirical selection probability of each predictor by applying a high-dimensional variable selection procedure, e.g., LASSO, to a different subset of half the observations for all λ values from a given set, and then retains only predictors with a selection probability above a user-chosen threshold. SS is implemented in `R/stabs` package version 0.6-3 ([Hofner and Hothorn 2017](#)) and can also be adapted to a multivariate framework. For QTL detection on experimental data, the probability threshold we applied was 0.6 for LASSO.SS and 0.7 for MTV_LASSO.SS.

Marginal false discovery rate:

Marginal false discovery rate (mFDR) has been defined by [Brehehy \(2019\)](#) as a modified version of the FDR in which those variables correlated with the causative features are not considered as false discovery. This study provided an accurate estimation of mFDR for a given λ when using EN or LASSO, thus allowing the selection of a more conservative value of λ in order to remain below a given mFDR threshold. We applied mFDR with the `R/nvcreg` package version 3.12.0 ([Brehehy 2019](#)). For QTL detection on experimental data, we set mFDR to 10% for LASSO.mFDR and EN.mFDR. To our knowledge, this approach had not been adapted yet to a multivariate framework.

Evaluation and comparison of methods

All methods were compared on two aspects: their ability to predict genotypic values, and their ability to select relevant markers, i.e., to detect QTLs. To assess the prediction of genotypic values on simulated data, we used the Pearson's correlation coefficient between predicted and simulated genotypic values (prediction accuracy). On experimental data, we used the same criterion, but the actual genotypic values being unknown, we used their empirical BLUPs instead (predictive ability).

For QTL detection on simulated data, the methods were compared using criteria of binary classification based on the numbers of true positives and false negatives. On experimental data, because true QTLs are unknown, no such comparison could be performed; instead, we compared the outcome of the different methods and QTLs were deemed reliable when found with several methods.

Genomic prediction

A nested cross-validation (CV) was applied to assess prediction by the various methods.

- An outer $k1$ - fold CV was performed to estimate the performance metrics, with an inner $k2$ - fold CV applied to the training set of each outer fold to find the optimal tuning parameters for the method under study ([Supplementary Figure S5](#)). Both $k1$ and $k2$ were set to 5 (see [Arlot and Lerasle \(2016\)](#)). The partitions of the outer CV were kept constant among traits and methods.
- For IM methods, the optimal tuning parameter was the LOD threshold obtained from permutations, and the effects for the four additive genotypic classes (ac, ad, bc, and bd) were estimated by fitting a multiple linear regression model with genotype probabilities at all QTL peak positions as predictors, using `R/stat::lm`. For penalized regression methods, parameters were optimized with specific functions such as `cv.glmnet` and `cv.spring`.
- As performance metrics, we used mainly Pearson's correlation (`corP`) but we also calculated the root mean square predicted error (RMSPE), Spearman correlation (`corS`), the model efficiency ([Mayer and Butler 1993](#)), and test statistics on bias and slope from the linear regression of observations on predictions ([Piñeiro et al. 2008](#)).

For experimental data, the whole nested cross-validation process was repeated 10 times ($r = 10$), whereas for simulated data it was performed only once, but on 10 different simulation replicates ($r = 1$ and $t = 10$). The 14 traits were analyzed jointly for MTV_RR, MTV_LASSO, and MTV_EN. But for SPRING, since analyzing all traits together was computationally too heavy, we split traits into three groups by hierarchical clustering ([Supplementary Figure S6](#)) performed with `R/hclust` applied to

Table 2 Parameter ranges for ROC curve computation, for comparing predictor selection performance of different methods

Method	SIM/MIM	LASSO/MTV_LASSO	Stability Selection	SPRING	EN	mFDR
Parameter name	LOD	λ	probability threshold	λ_1	λ	mFDR
Lowest constraint	0	10e-5	0.5	10e-8	10e-4	0.3
Highest constraint	14	0.25	0.9	0.25	8	0

genotypic BLUPs. All traits within a given cluster were analyzed together.

For simulated data with the same heritability values for both traits, performance results were averaged not only over simulation replicates and partitions of outer CV, but also over traits, because both traits were equivalent in terms of simulation parameters. For simulated data with different heritability values, performance results were averaged only over simulation replicates and partitions of outer CV. For experimental data, performance results were averaged over partitions of outer CV and outer CV repetitions.

QTL detection

Simulated data:

The quality of a predictor selection method is usually assessed through the relationship between statistical power (i.e., the true positive rate, TPR) and type I error rate (i.e. the false positive rate, FPR). To compare methods, we thus used the ROC (receiver operating characteristic) curve (Swets et al. 1979), which is the plot of TPR as a function of FPR over a range of parameters (Table 2), and the pAUC (partial Area Under the Curve; McClish 1989; Dodd and Pepe 2003). Any marker selected at ± 2 cM of a simulated QTL was counted as a True Positive.

For methods with two tuning parameters, one parameter was kept constant (α at 0.7 for EN and EN.mFDR, and λ_2 at 10e-8 for SPRING). We tested several α values for EN but it did not change much the results (data not shown). For MIM, the maximum number of QTLs that can be integrated into the model was set to 10.

Experimental data:

Comparison between methods was based on the number of detected QTLs, the magnitude of their effects, and the percentage of variance globally explained by all detected QTLs.

For MTV_LASSO and SPRING, we split traits into three groups as described above, for computational reasons (for SPRING) and to test whether such splitting evidenced more reliable QTLs (for MTV_LASSO). The parameters of penalized methods were tuned by cross-validation, with MSE as the cost function. We compared predictor selection between methods in terms of the number of common selected markers per trait, i.e., the intersection between markers selected by penalized methods and markers inside confidence intervals found by IM methods. Then all markers in high LD with those selected were considered as selected too. The threshold was defined as the 95% quantile of LD value distribution, for all pairs of markers belonging to the same chromosome (Supplementary Figure S7), which gave a LD threshold of 0.84.

We deemed selected markers as highly reliable if they were either (i) selected by at least five methods, whatever the methods, (ii) or selected by both EN.mFDR and MIM (see Results). Then, we defined a highly reliable QTL as the interval of ± 3 cM around each highly reliable marker (Price 2006; Viana et al. 2016b), as predicted by polynomial local regression (loess) fitting of genetic positions to physical position. When several markers were selected inside the 6 cM interval, the QTL interval was extended

accordingly. The genetic positions of this interval were then converted into physical positions, by fitting loess. QTLs overlapping for several traits on the SNP map were merged into a single QTL, by physical intervals' union.

Candidate genes exploration:

After merging the most highly reliable QTLs colocalized between traits, we searched for underlying candidate genes. We retrieved the list of genes overlapping the intervals of our QTLs from the reference *Vitis* genome 12X.v2 and the VCost.v3 annotation (Canaguier et al. 2017). We then used the correspondence between IGGP (International Grapevine Genome Program) and NCBI RefSeq gene model identifiers provided by URGI (<https://urgi.versailles.inra.fr/Species/Vitis/Annotations>, last accessed on 07-21-21) to identify putative functions from NCBI, when available. For those genes with a putative function, we then refined the analysis to retrieve additional information about their function and expression. We searched UniProt (<https://www.uniprot.org/>, last accessed on 07-21-21) and TAIR (<https://www.arabidopsis.org/>, last accessed on 07-21-21) databases based on homologies to access a complete description of gene function, name, and corresponding locus in *Arabidopsis*. In addition, we used the GREAT (GRape Expression Atlas) RNA-seq data analysis workflow (<https://great.colmar.inrae.fr/app/GREAT>, last accessed on 07-21-21), which gathers published expression data, to assess the level of expression of our candidate genes in grapevine leaves and shoots, relevant organs for the traits under study. RNA-seq data are normalized as detailed in the 'User manual' section of the GREAT platform: "from the raw read counts, the normalized counts (library size normalization) and the RPKM (gene size normalization) are calculated for each gene in each sample." Data were retrieved with all filters set to "Select All" except for the organ considered that was restricted to 'Leaves' and 'Shoot'.

Results

Comparison of methods with simulated data prediction: cross-validation results

Traits with the same heritability value:

Methods were compared for prediction accuracy by applying cross-validation on simulated data with four different configurations and four heritability values.

Mean Pearson's correlation coefficient varied from 0.16 to 0.98, with a strong effect of heritability on prediction accuracy in all configurations, for the seven main methods (Figure 1). As expected, MIM performed very well in the "major" configurations across all heritability values but yielded the least accurate prediction in the "minor" ones. On the opposite, RR performed very well in the "minor" configurations, but yielded the least accurate prediction in the "major" ones. EN prediction performance was always intermediate between those of RR and LASSO. QTL distribution among traits – "same" (for QTLs at the same positions) or "diff" (for QTLs at different positions) - had very little effect on prediction accuracy. Moreover, we did not observe any superiority of multivariate methods over univariate ones, despite

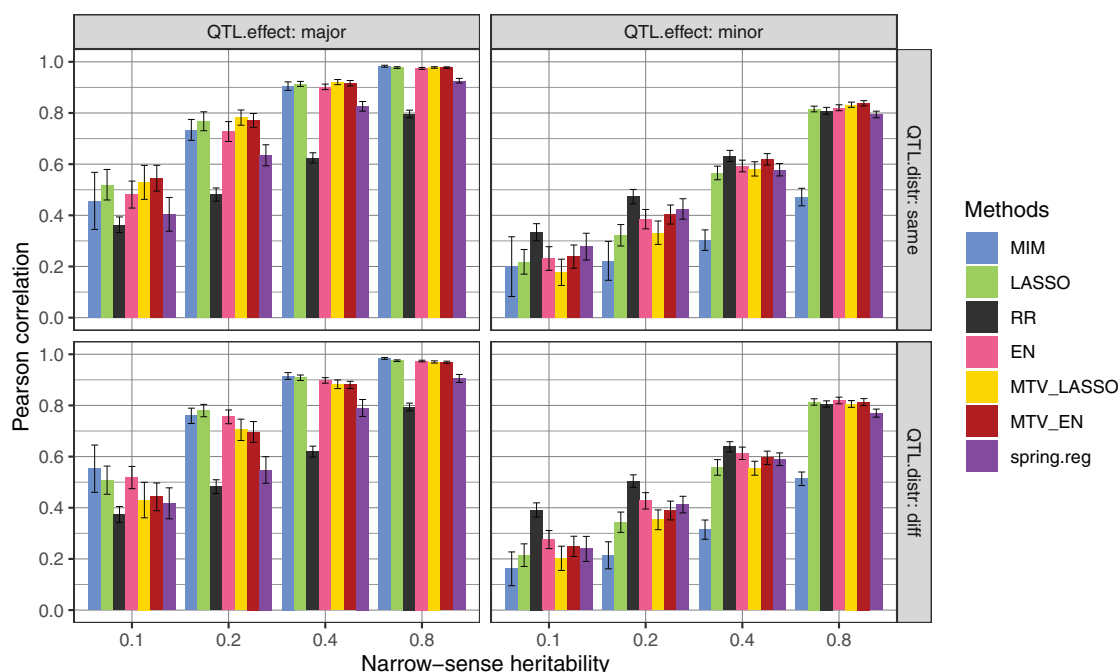


Figure 1 Genomic prediction accuracy (Pearson’s correlation between predicted and true genotypic values) of seven methods applied to 3961 markers and two simulated traits in a bi-parental population with different heritability values and four QTL configurations (number \times distribution among traits). major: 2 QTLs; minor: 50 QTLs; same: QTLs at the same positions for both traits; diff: QTLs at different positions between traits. For each heritability value and configuration, prediction accuracy was averaged over 100 values (2 traits \times 10 simulation replicates \times 5 cross-validation folds). The error bar corresponds to the 95% confidence interval around the mean.

the strong genetic correlation simulated between traits ($\rho_B=0.8$) and no correlation between errors.

The prediction accuracy of four additional methods is shown in [Supplementary Figure S8](#) and prediction accuracy values. Other performance metrics (see Materials and Methods) are given for all methods in [Supplementary Table S9](#). All IM methods yielded equivalent prediction accuracy. LASSO.GB did not improve performance compared to LASSO. MTV_RR showed equivalent performance to univariate RR. Prediction accuracy with *spring.dir.ols* was always lower than with *spring.reg*, and even very low for “minor” configurations. With 100 or 1000 simulated QTLs (under each QTL distribution) the ranking of methods based on prediction accuracy did not change compared to “minor” configurations ([Supplementary Figure S10](#)).

Traits with different heritability values:

To further compare prediction accuracy between univariate and multivariate methods, we simulated two correlated traits with different heritability values, 0.1 and 0.5. MTV_LASSO performed slightly better than univariate LASSO for the lowest heritability trait; however, differences were not significant ([Supplementary Figure S11](#)). On the opposite, prediction accuracy was lower with MTV_LASSO than with univariate LASSO for the highest heritability trait, reaching quite low values with 200 simulated QTLs. The same trends were also visible for MTV_EN and EN. MTV_RR never improved prediction over RR and *spring.reg* never performed better than RR.

Since these results were unexpected, we also compared prediction accuracy of the above methods using the simulated data published by [Jia and Jannink \(2012\)](#). We obtained very similar differences among methods to those with our own simulated data, even though prediction accuracy was higher in all cases ([Supplementary Figure S12](#)).

QTL detection: ROC curve results

We compared the main methods mentioned above (except RR that does not perform marker selection), as well as some robust extensions, for their marker selection performance, by means of ROC curves, using the same simulated data in the four configurations ([Figure 2](#)). The closer a ROC curve obtained through a given method approaches the optimum point (i.e., FPR = 0 and TPR = 1), the better is the method’s selection performance. As expected, IM methods (SIM and MIM) showed low selection performance when many minor QTLs were simulated and high selection performance when a few major QTLs were simulated. Note that the MIM curve was hardly visible; it roughly overlapped with the SIM curve but stopped at a low FPR because it could not select many QTLs by design.

The penalized regression methods always performed at least as well as the IM methods and even much better in the case of “minor” configurations. Among penalized methods, none was clearly better than the others in all configurations, except for a slight superiority of MTV_LASSO in the “same_minor” configuration. These methods, and particularly *spring.dir.ols*, displayed high variability in classification results with two simulated QTLs (“major” configurations). Indeed, when one QTL was not detected in one trait, impact on TPR was stronger than with 50 simulated QTLs.

The most interesting part of the ROC curve for QTL detection is the left most part, i.e., that with a low FPR (e.g., below 0.1). We thus calculated the partial Area Under the Curve (pAUC) for FPR between 0 and 0.1 for methods reaching that value ([Supplementary Figure S13](#)). EN resulted in constantly high pAUC across configurations and heritability values. In contrast, pAUC for SIM was quite high at low heritability values for the “same_major” configuration but dropped for other configurations and heritability values.

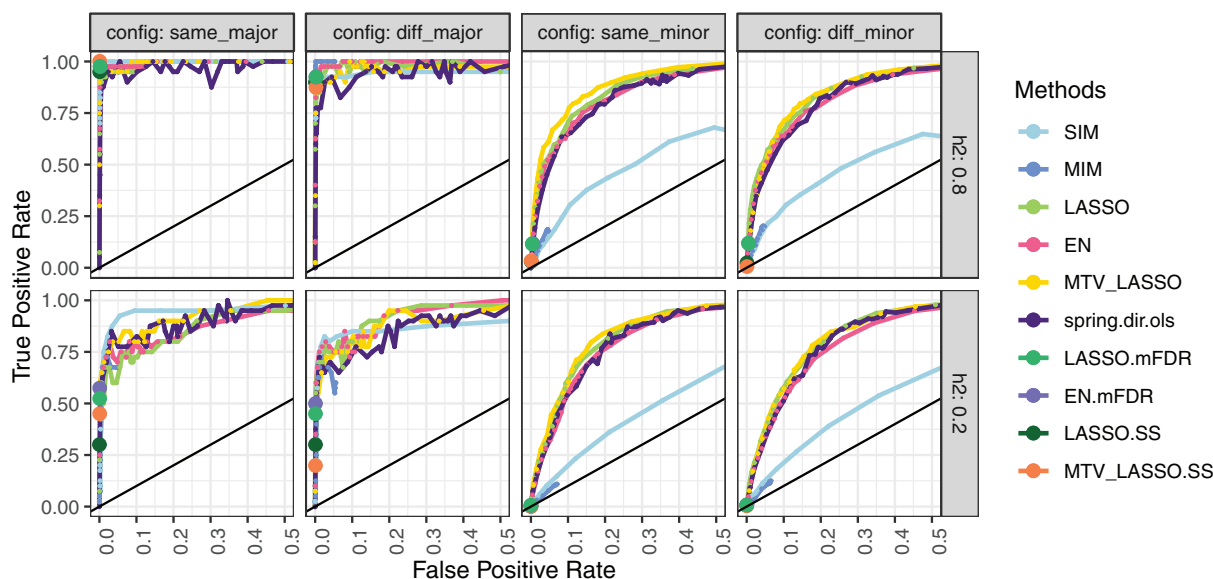


Figure 2 ROC curves for 10 methods applied to 3961 markers and two simulated traits in a bi-parental population with two heritability values and four QTL configurations (number \times distribution among traits). major: 2 QTLs; minor: 50 QTLs; same: QTLs at the same positions for both traits; diff: QTLs at different positions between traits. Results are averaged over 2 traits \times 10 simulation replicates. TPR: (number of correctly found QTLs/number of simulated QTLs), FPR: (number of falsely found QTLs/number of markers outside a QTL). For robust methods (mFDR and SS), as the FPR remained very low, we display only a single point corresponding to the lowest parameter constraint and thus to the highest TPR.

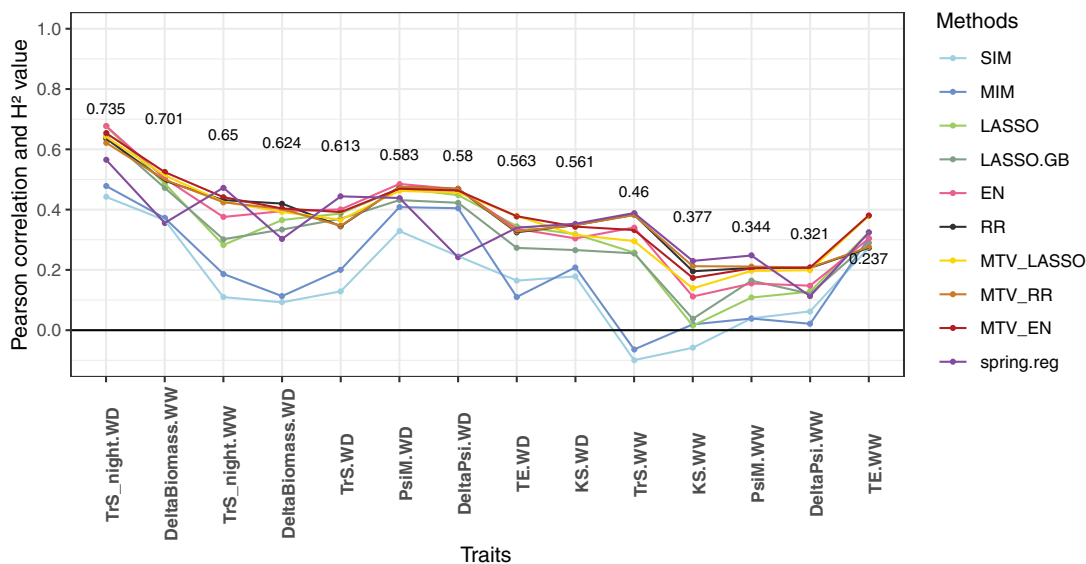


Figure 3 Mean genomic predictive ability (Pearson's correlation between genotypic BLUPs and their predicted values), obtained by cross-validation for 10 methods applied to 14 traits related to water deficit and GBS gene-dose data, within a grapevine bi-parental population. Broad-sense heritability values are reported for each trait (y-position of the number corresponds to heritability estimate). Traits are ordered by decreasing heritability. For each trait, predictive ability is averaged over 10 cross-validation replicates \times 5 cross-validation folds).

Results on experimental data

Genomic predictive ability:

Mean genomic predictive ability per trait ranged from -0.10 to 0.68 (Figure 3 and Supplementary Table S14). It decreased with broad-sense heritability. IM methods (in blue) were always among the three poorest methods for prediction. Based on the mean predictive ability averaged across traits, MTV_EN yielded the highest correlation (0.384), followed by RR (0.3721), MTV_RR (0.3716), MTV_LASSO (0.369), EN (0.357), *spring.reg* (0.344), LASSO

(0.329), LASSO.GB (0.313), MIM (0.200), and SIM (0.162). However, based on the number of traits for which each method gave the best prediction, *spring.reg* had the highest score, with 6 traits out of 14, followed by MTV_EN (3 out of 14) and EN (2 out of 14).

In a nutshell, MTV_EN and RR, tied with MTV_RR, provided the best mean predictive ability across traits. Even though *spring.reg* outperformed them for some traits, its performance was unstable and especially low for *DeltaBiomass.WW*, *DeltaBiomass.WD*, *DeltaPsi.WW*, and *DeltaPsi.WD*. For computational reasons, all

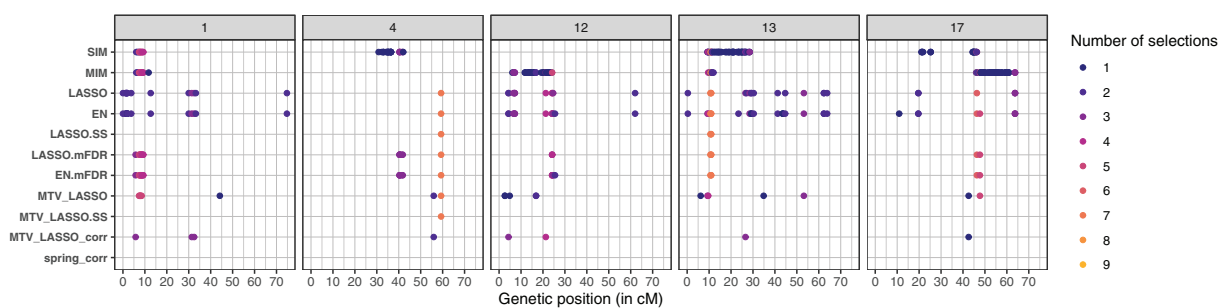


Figure 4 Marker selection by all methods for *TrS_night.WD* trait on chromosomes 1, 4, 12, 13, and 17. Each marker selected by a given method is represented by a colored point, the color indicating the number of methods that have selected that specific marker. The boxes correspond to chromosomes and the x-axis to the position along the genetic map (in cM).

traits could not be analyzed together with *spring.reg*, but were divided into three groups. These four traits with low predictive ability belonged to the same group. Yet, the effect of group membership on predictive ability was not significant at 5% (P -value = 0.30 and percentage of variance explained = 24%).

QTL detection:

To address the intersection of SNP selection by all methods, and determine the number of reliable intervals (QTLs) and their position, we examined in detail marker selection for each trait and chromosome. These results are given in [Supplementary Table S15](#), together with genetic and physical positions and the percentage of variance explained. These results are plotted in [Figure 4](#) for night-time transpiration under water deficit (*TrS_night.WD*) and in [Supplementary Figure S16](#) for all traits.

Most of the time, more markers were selected for traits under water deficit than for traits in well-watered conditions, and they were most often selected by several methods. Penalized methods tended to select exactly the same markers, not only close ones; for example, for *TrS_night.WD* on chromosome 4, the same marker (at physical position 21,079,664 bp) was selected by seven different methods ([Figure 4](#)).

We considered markers selected by both MIM and EN.mFDR as highly reliable ones for three reasons: (1) markers selected by both MIM and EN were considered as reliable ones, because most markers selected by LASSO were also selected by EN, whereas MIM marker selection was quite different; (2) simulations showed that MIM and mFDR methods led to a very low FPR; (3) these methods belong to different method classes (IM vs penalized regression). We also considered as highly reliable those markers selected by at least five methods. These criteria resulted in a set of 59 highly reliable selected markers, which were converted to genetic intervals of ± 3 cM around each selected marker. Overlapping intervals per trait were merged, resulting in 25 highly reliable QTLs.

These 25 QTLs involved nine traits, mostly under water deficit, and were located on seven chromosomes ([Supplementary Figure S17](#), [Supplementary Table S18](#)). Some QTLs colocalized for different traits, such as on chromosome 1, and had similar distributions of genotypic BLUPs according to genotypic classes ([Supplementary Figure S19](#)).

Among these 25 QTLs, we found eight new highly reliable QTLs compared to [Coupel-Ledru et al. \(2014, 2016\)](#), among which five were not detected by MIM. In particular, a completely new QTL for *TrS_night.WD* was found alone on chromosome 12. Most other new QTLs were colocalized with QTLs previously found in single-year analysis and/or for the other watering condition.

Notably, we observed colocalization of *TrS_night.WD*, *TE.WD* and *DeltaBiomass.WD* QTLs on chromosomes 4 and 17.

In total, the percentage of variance explained (adjusted R^2) per trait was 51.3% for *TrS_night.WD* (36% in 2012 for [Coupel-Ledru et al. 2016](#)), 33.9% for *PsiM.WD*, 31.4% for *DeltaPsi.WD*, 26.9% for *DeltaBiomass.WD*, 19.4% for *TE.WD*, 18.6% for *TE.WD*, 17.0% for *KS.WD*, 14.9% for *DeltaBiomass.WD*, and 8.5% for *TrS.WD*.

Candidate genes

After merging the QTLs colocalized between traits, we obtained 12 intervals, located on chromosomes 1, 4, 10, 12, 13, 17, and 18, harboring a total of 3461 genes according to the VCost.v3 annotation ([Canaguier et al. 2017](#)). Among them, 2379 had an NCBI Refseq identifier and 1757 a putative function ([Supplementary Table S20](#)). We then focused our analysis on the eight “new” intervals, i.e., those that were not overlapping with QTL intervals repeated over years by [Coupel-Ledru et al. \(2014, 2016\)](#). They encompassed 1155 genes, half of which were annotated. We were able to retrieve from TAIR and/or UniProt a more precise description of the gene function for 86% of the annotated genes ([Supplementary Table S20](#)). The remaining ones either did not have any homologous gene in *Arabidopsis* or were not described in the above-mentioned databases. RNA-seq data was available on the GREAT platform for 90% of the annotated genes. We further focused on the highly reliable QTL co-localized on chromosome 4 for *TE*, *TrS_night* and *DeltaBiomass* under various conditions. We proceeded to a functional classification of the 161 annotated genes underlying this QTL, based on the full description previously retrieved ([Supplementary Tables S5](#) and [S21](#)). For 75 genes, an integrated function at the plant or organ level was explicitly quoted in the homologous-based description. We grouped these integrated functions into 12 major groups ([Figure 5](#)). For a substantial part of genes, functions consistently related to traits involved: 15 genes related to hydraulics (stomata, xylem, and trichomes), relevant for *TrS_night* and thus *TE*; 27 to growth or development and one to photosynthesis, both relevant to *DeltaBiomass* and thus *TE*. For the 86 genes for which no integrated function was explicitly quoted, we further built a classification based on their cellular or molecular function. Among them, we found six genes related to carbon metabolism, one to wall formation (both relevant for *DeltaBiomass*) and six to drought stress signaling and drought-related hormones (relevant for *TrS_night*).

Discussion

This study contributes to our knowledge of the complex genetic determinism of vegetative traits under different watering

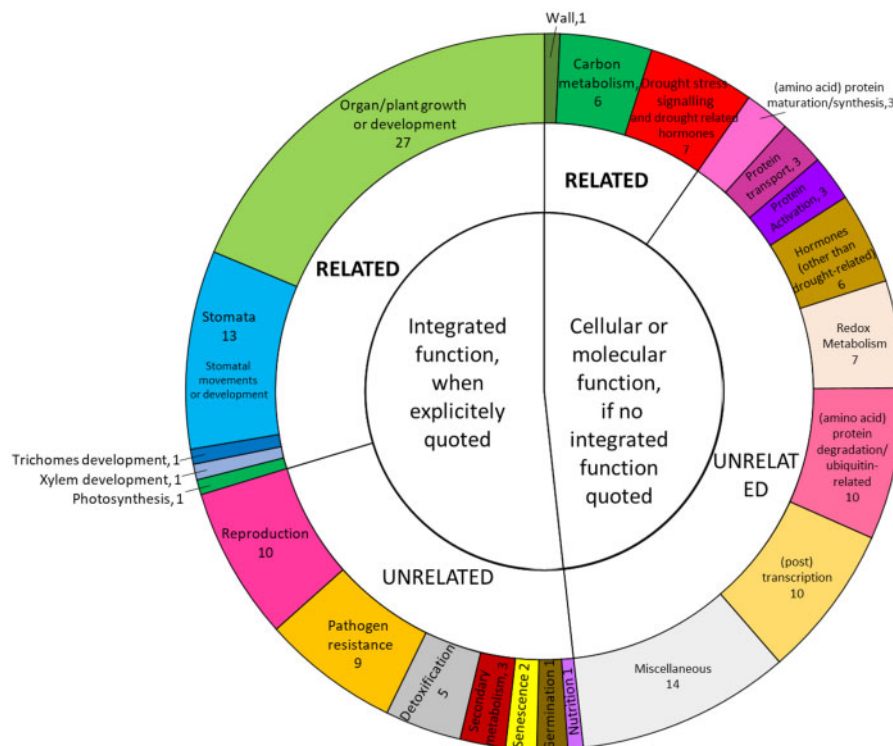


Figure 5 Functional classification of the annotated genes underlying the highly reliable QTL detected on chromosome 4 for night-time transpiration, growth, and TE. Hierarchical classification of the 161 genes based on their functions. See [Supplementary Table S21](#) for the details of this classification. When an integrated function at the organ or plant level was explicitly quoted in the gene annotation, genes were classified on this basis. When no integrated function was explicitly quoted, they were classified based on their cellular or molecular function. In both cases, functions were then classified as “Related” if related to the traits of interest in this QTL, or “Unrelated” if not.

conditions in three different ways. We compared by simulation several univariate and multivariate methods for genomic prediction and QTL detection, and re-analyzed grapevine phenotypes obtained under semi-controlled conditions. In particular, we showed that penalized methods are valuable not only for prediction but also for QTL detection. Indeed, we found new QTLs using these methods and identified relevant candidate genes.

Methodological aspects: method comparison

Handling linkage disequilibrium:

IM methods estimate genotypic probabilities between markers according to a genetic map which is computationally costly to build. On the other hand, most penalized methods do not require any previous knowledge on LD.

The LASSO assumption that all predictor variables are independent is all the more violated that there are many markers. In the case of a group of correlated predictors (*e.g.*, SNPs in LD), EN selects either no or all predictors within the group with close estimated effects (Zou and Hastie 2005), whereas LASSO selects a single predictor. In that sense, EN aims at correcting the drawbacks of LASSO when predictor variables are highly correlated. By exploring a large number of configurations of the finite-sample high-dimensional regression problem, Wang *et al.* (2020) showed that EN is competitive for both prediction and selection in most cases with highly correlated predictors. In agreement with these results, we showed that EN performed well for both prediction and selection on our simulated data, and that multivariate EN performed best for prediction on grapevine experimental data.

It would be interesting to test whether EN still remains the main default method when applied to a population with a shorter LD, *e.g.*, a diversity panel as defined in Nicolas *et al.* (2016).

Indeed, the ranking of methods is likely to depend not only on linkage disequilibrium and population size, but also on the genetic architecture of the traits of interest as well as the accuracy with which phenotypic values were obtained, and all these variables can interact with each other, but studying this was out of the scope of the current work.

Comparison between interval-mapping and penalized regression methods for genomic prediction:

As expected, IM methods performed poorly to predict accurate genotypic values when the QTL number was large. These conclusions are in agreement with previous studies (Figure 1 and Supplementary Figure S8), even though most implemented marker selection methods other than interval-mapping (Bernardo and Yu 2007; Lorenzana and Bernardo 2009; Mayor and Bernardo 2009; Olatoye *et al.* 2019). This confirms that for complex traits, genomic prediction should not be based only on QTLs detected by IM methods.

None of the penalized univariate methods performed optimally in all cases (Figures 1 and 3 and Supplementary Figure S8), as also found in the literature (Heslot *et al.* 2012; Riedelshheimer *et al.* 2012; Azodi *et al.* 2019). As shown by simulation, RR was better adapted to highly polygenic genetic architecture, whereas LASSO was better adapted to a few major QTLs. Moreover, in the case of many minor QTLs, RR was the most stable method across heritability values, as previously described for several traits and species (Heslot *et al.* 2012; Azodi *et al.* 2019). However, RR prediction accuracy dropped when the QTL number was too small, whereas EN still predicted as well as LASSO. EN was hence well adapted to various numbers and distributions of QTLs.

Multivariate vs univariate:

When the same heritability was simulated for both trait variables, no superiority of multivariate methods was observed, even when both traits had QTLs at the same positions (Figure 1 and Supplementary Figure S8). When different heritability values were simulated for the two traits, we observed a slight superiority of MTV_LASSO (resp. MTV_EN) over LASSO (resp. EN) only in the “same” and “major” configuration (with both traits sharing the same two QTLs) for the trait with small heritability (Supplementary Figure S11).

Other authors who tested multivariate GP on simulated data systematically applied different heritability values; they found a superiority of multivariate methods over univariate ones for the trait with the smallest heritability (Calus and Veerkamp 2011; Guo et al. 2014; Jiang et al. 2015; Dagnachew and Meuwissen 2019). However, all these studies were based on a smaller, more favorable, p/n ratio, a key component of high-dimensional models (Verzelen 2012). For example, in Jia and Jannink (2012), their 500 observations for 2020 predictors correspond to a ratio of ~ 4 , compared to our 188 observations for 3961 predictors corresponding to a ratio of ~ 21 . Indeed, parameters n and p are involved in the sample complexity function defined in Obozinski et al. (2011), which predicts the theoretical cases where MTV_LASSO is superior to its univariate counterpart in terms of variable selection. Accordingly, applying our methods to Jia and Jannink (2012) data allowed us to evidence a larger difference between univariate and multivariate LASSO than with our simulated data.

Unexpectedly, when reanalyzing the data simulated by Jia and Jannink (2012), we obtained lower prediction accuracy with our MTV_LASSO (Supplementary Figure S11) than they did with their multivariate BayesA (their Figure 1A). A similar result in a univariate setting was found by Guan and Stephens (2011), who compared BSVR (comparable to BayesA) and the LASSO. They found that BSVR had markedly higher power than LASSO. Moreover, the parameters of both BSVR [in Guan and Stephens (2011)] and BayesA [in Jia and Jannink (2012)] were estimated with a MCMC algorithm. No inner cross-validation was needed, hence the sample used to train the model was larger. This difference may explain why Figure 1A from Jia and Jannink (2012) shows better prediction accuracies for multi-trait models compared to their single-trait counterparts, although no confidence interval was displayed. Note that our RR prediction accuracies were close to those of their GBLUP (univariate and multivariate). In conclusion, prediction accuracy is affected both by the dimension of the problem (i.e., n and p) and the method used (i.e., Bayesian with MCMC or cross-validation).

For experimental data, we observed that MTV_LASSO (respectively MTV_EN) was superior to LASSO (resp. EN) for the five traits with the smallest heritability (Figure 3). Such this improvement suggests that MTV_LASSO (resp. MTV_EN) was able to borrow signals from the most heritable traits to the least heritable ones, likely because of a partially overlapping genetic architecture between these traits. This interpretation is reinforced by the fact that a QTL for low-H2 trait, *TE.WW*, colocalizes on chromosome 4 with QTLs for four high-to-moderate-H2 traits (*TrS_night.WD*, *DeltaBiomass.WW*, *DeltaBiomass.WD* and *TE.WD*). This improvement was not found in Jia and Jannink (2012), who also tested their methods on real pine data from Resende et al. (2012). These observations suggest that the number of traits analyzed (14 in our case and 2 in Jia and Jannink 2012 study) may also play a role in the gain in prediction accuracy of multivariate over univariate methods.

Furthermore, we simulated data with various levels of residual correlation among traits (0, 0.4, and 1) but this did not significantly change prediction results (data not shown). A more detailed methodological analysis is out of the scope of the current work.

Comparison between interval-mapping and penalized regression methods for QTL detection:

Comparison with the ROC curve between IM and penalized regression methods for marker selection has not been extensively studied before. As expected, we found that IM methods are well adapted to detect a few major QTLs but not many minor QTLs (Figure 2). Yi et al. (2015) similarly compared the FDR and TPR reached by single marker analysis and different penalized regression methods, some of which being adapted to control FDR; they found contrasting results, depending on the criteria studied (modified version of TPR or FDR). However, they focused only on an association panel whereas we worked on a bi-parental population. Other authors (Cho et al. 2010; Li and Sillanpää 2012; Waldmann et al. 2013) successfully applied LASSO or EN for performing GWAS, but without comparing IM and penalized methods for QTL detection. Moreover, we found that penalized methods could be as good at marker selection as IM methods, and even far better when there were many minor QTLs. Among the penalized methods we compared, none clearly outperformed the others for marker selection in all configurations.

Multivariate vs univariate:

As MTV_LASSO selects one predictor for all traits, its superiority over univariate LASSO depends on QTL distribution across traits, notably on the amount of genetic basis shared by the traits (Obozinski et al. 2011). However, as for prediction, we showed that MTV_LASSO performance was not different whether QTLs were at the same or at different positions across traits (Figure 2). Nevertheless, we observed that MTV_LASSO was slightly better than LASSO when many QTLs were simulated.

SPRING had never been evaluated before for its quality of predictor selection. As for prediction, SPRING showed unstable results across our simulation replicates and hyper-parameter values. However, for the ROC curve, we did not include predictor structure in the model, which may have hampered marker selection quality.

Efficient default method for both QTL detection and genomic prediction:

IM methods were designed for marker selection; hence they are not expected to be optimal for prediction, as confirmed in our study. Among penalized regression methods, some may be better at prediction than at marker selection, and vice versa. For example, our results showed that EN performed well across several configurations for both aims. Some methods such as SPRING are specially adapted to handle both purposes but this method produced too variable results for either prediction or QTL detection. However, SPRING is a recent method that still can be improved to correct this drawback.

New penalized regression methods are continuously being developed. In particular, graph structured sparse subset selection (Grass) recently proved to outperform existing methods for both prediction and predictor selection, thanks to a L0 regularization that limits the number of nonzero coefficients in the model (Do et al. 2020). It could be tested on our data when available. Moreover, multivariate methods are presented as being more

efficient at using the whole signal in the data, whether for marker selection (Inouye et al. 2012) or prediction (Jia and Jannink 2012; Guo et al. 2014), but our results revealed no systematic advantage of multivariate methods over univariate ones for both aims.

Using penalized methods for both marker selection and genomic prediction requires adapted hyper-parameter values. For EN, LASSO, and SPRING, the λ value controls sparsity (e.g., the number of selected markers). Thus, the optimal λ value might not be the same if the aim is to limit FPR or to maximize predictive ability (Li and Sillanpää 2012). For prediction, we traditionally use cross-validation to tune hyper-parameters by minimizing MSE. For marker selection, there is no direct equivalence. That is why we tested extensions of these methods (mFDR and SS) that control sparsity for robust marker selection; both proved efficient in selecting the most relevant markers.

In order to shed light on the link between prediction accuracy and marker selection, we plotted the prediction accuracy at each point of the ROC curve for EN and EN.mFDR against FPR for minor configurations (with 50 simulated QTLs) (Supplementary Figure S22). For EN, we showed that prediction accuracy reached its maximum when FPR was below 0.05. Then, FPR increased while prediction accuracy decreased until it reached a plateau. This means that prediction quality is intimately linked to selection quality, especially at low heritability. For EN.mFDR, FPR always stayed below 0.015 but prediction accuracy was lower.

As a consequence, as an efficient default method, we advise at this stage to apply EN for performing genomic prediction, and its extension EN.mFDR for performing sparser marker selection.

Genetic determinism and prediction of grapevine response to water deficit:

Based on experimental data on the Syrah \times Grenache progeny (new genotypic data and already published phenotypic data), we compared the same methods as above for both prediction and marker selection. To the best of our knowledge, grapevine GP within a bi-parental family has only been applied to a limited number of traits, with very few methods and never multivariate GP. Fodor et al. (2014) studied GP in grapevine with simulated data on a diverse structured population; they tested RR-BLUP, Bayesian Lasso, and a combination of marker selection and RR. Viana et al. (2016a) used an inter-specific grapevine bi-parental population. They predicted cluster and berry phenotypes (number and length of clusters, number, and weight of berries, juice pH, titrable acidity) with RR-BLUP and Bayesian LASSO applied to table grape breeding. In addition to yielding further insights into method comparison beyond those obtained by simulation, our study brought valuable novel biological knowledge about grapevine water use under different watering conditions. Indeed, new methods and the new SNP genetic map allowed us to find novel QTLs, as compared to those previously detected with the same phenotypic data (Coupel-Ledru et al. 2014, 2016). Our study also provides novel results of practical interest to grapevine breeders. We showed what predictive ability they can expect for drought-related traits within a progeny: here, always higher than 0.3, and up to more than 0.65 for some traits. Even though these traits are difficult to phenotype, they correspond to crucial breeding targets in the context of climate change. Our results may help motivate their phenotyping in the training panels of breeding programs.

Predictive ability and genetic architecture:

Among univariate penalized methods, RR generally had equivalent or better predictive ability compared to LASSO. For the traits

with the largest discrepancy between RR and LASSO, this suggests that trait variability was rather due to many minor QTLs than to a few major ones. On the other hand, for a few traits, e.g., PsiM.WD, DeltaPsi.WD, and TE.WW, predictive abilities with sparse methods (s, LASSO, and IM methods) were better than with RR, suggesting a genetic architecture with few major QTLs rather than an infinitesimal one in those cases.

Finally, while not considered by the penalized methods used, nonadditive genetic effects such as epistasis could be involved. We, therefore, tested the superiority of LASSO.GB over LASSO. Extreme Gradient Boosting methods are indeed among the best machine learning methods (Chen and Guestrin 2016). LASSO.GB did not markedly increase predictive ability on experimental data (Figure 3). However, we cannot exclude that this might be due to a poor optimization of Extreme Gradient Boosting parameters or to an insufficient number of observations to correctly fit the model. We also tested if coding differently the design matrix to estimate dominance genetic effects improved predictive ability and it was not the case (data not shown).

Candidate gene analysis:

The thorough methodology deployed for candidate genes analysis allowed us not only to retrieve a list of genes potentially underlying the QTLs of interest, but also to classify them based on their function and expression to point at the most likely candidates. We focused on the highly reliable QTL detected on chromosome 4 for TrS_night, TE, and DeltaBiomass. TrS_night QTL was previously described as a promising target for marker-assisted selection, as alleles limiting night-time transpiration also favor plant growth, resulting in a doubly, beneficial impact on improving TE (Coupel-Ledru et al. 2016). Moreover, this QTL was found using seven methods. Among the plethora of integrated functions represented within the list of annotated genes underlying this QTL, we show here that a subset of more likely candidates can be defined as possibly related to the traits of interest. On the one hand, these include genes related to broad-sense hydraulics and water loss, with a possible direct impact on TrS_night: seven genes involved in stomatal development, nine involved in stomatal opening—sometimes through the abscisic acid signaling pathway—, one related to xylem development and one to trichome development (Supplementary Table S21). One of these genes, the trihelix transcription factor GT-2 (Vitvi04g01604), was specifically shown to impact transpiration and TE in *Arabidopsis* by acting as a negative regulator of stomatal density. On the other hand, 27 genes in the list are directly related to growth, development, or photosynthesis, hinting to a possible direct impact on DeltaBiomass. A histidine kinase 1 (Vitvi04g01483) may be a particularly interesting candidate for its multiple roles in *Arabidopsis* in ABA signaling, stomatal development, and plant growth, hence potentially simultaneously acting on both components of TE. Both these likely candidates were often highly expressed in grapevine leaves. More precise analyses of these candidate genes, including functional genomics work and possibly gene editing, will now be necessary to identify the causative polymorphisms under these new QTLs.

Conclusions

Rather than decoupling genomic prediction from the identification of major QTLs, we argue for the need to pursue both goals jointly. Indeed, they provide complementary information on the genetic architecture of the target traits, as well the key

underlying functions. Our study provided encouraging findings for further implementing genomic prediction in grapevine breeding programs. Applied to both simulated and 14 experimental traits, univariate and multivariate Elastic Net proved to be efficient for both goals, followed by mFDR control for the robust identification of QTLs. Moreover, of interest to plant biologists seeking to understand the response to water stress, our results highlighted several candidate genes underlying integrated traits such as night-time transpiration, TE, and biomass production. For some, their putative functions suggest causal links with stomatal functioning, trichome development, or the ABA pathway.

Code availability

All software we used was free and open-source and most analyses were done with R (R Core Team 2020), notably graphs that were created using the ggplot2 package (Wickham 2016), version 3.3.2. All R scripts used for analysis, i.e. genetic mapping, simulation, phenotypic analysis, prediction, and QTL detection, are available online at <https://doi.org/10.15454/NOUQY2>. Many of the custom functions we used are available in a package for reproducibility purposes, R/rutilstimflutre (Flutre 2019).

Author contribution statement

T.F., A.D., and L.L.C. conceived the idea of the study and contributed to funding acquisition; A.C.L. and T.S. obtained the phenotypic data used in this work; C.B. and A.D. carried out the analysis of the genotypic data, with input data previously analyzed by T.F.; T.F. and J.C. performed preliminary multivariate genomic prediction analysis; A.C.L. and T.S. interpreted the results from the candidate gene analysis and wrote the corresponding parts in the manuscript; PT is the PhD supervisor of CB; CB wrote the original draft, which was reviewed and edited by all authors. All authors read and approved the final manuscript.

Ethical standards

The authors declare that the experiments comply with the current laws of the country in which they were carried out.

Data availability

Genotypic data, as well as the genetic map, are available online at <https://doi.org/10.15454/QEDX2V>. Raw phenotypic data are available in a second online repository at <https://doi.org/10.15454/YTRKV6>.

Supplementary material is available at G3 online.

Acknowledgments

We thank Marie Perrot-Dockès for her help concerning penalized regression methods, the South Green platform for computing facilities, notably Bertrand Pitollat, the GREAT platform, notably Amandine Velt and Camille Rustenholz, and Philippe Châtelet for his useful suggestions on the text.

Funding

We acknowledge the funding of the SelVi and FruitSelGen projects (BAP department and Selgen metaprogram of INRAE). Partial

funding of the PhD was provided by ANRT (Association nationale recherche technologie, grant number 2018/0577), IFV and Inter-Rhône.

Conflicts of interest

None declared.

Literature cited

- Adam-Blondon A-F, Bernole A, Faes G, Lamoureux D, Pateyron S, et al. 2005. Construction and characterization of BAC libraries from major grapevine cultivars. *Theor Appl Genet.* 110:1363–1371.
- Arlot S, Lerasle M. 2016. Choice of V for V-fold cross-validation in least-squares density estimation. *J Mach Learn Res.* 17:7256–7305.
- Arruda MP, Lipka AE, Brown PJ, Krill AM, Thurber C, et al. 2016. Comparing genomic selection and marker-assisted selection for Fusarium head blight resistance in wheat (*Triticum aestivum* L.). *Mol Breed.* 36:84.
- Azodi CB, Bolger E, McCarren A, Roantree M, Campos G. D L, et al. 2019. Benchmarking parametric and machine learning models for genomic prediction of complex traits. G3 (Bethesda). 9: 3691–3702.
- Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. *arXiv.* 1406:5823.[stat] arXiv: 1406.5823.
- Bernardo R. 2008. Molecular markers and selection for complex traits in plants: learning from the last 20 years. *Crop Sci.* 48: 1649–1664.
- Bernardo R, Yu J. 2007. Prospects for genomewide selection for quantitative traits in maize. *Crop Sci.* 47:1082–1090.
- Brehehy PJ. 2019. Marginal false discovery rates for penalized regression models. *Biostatistics.* 20:299–314.
- Broman KW, Wu H, Sen C, Churchill GA. 2003. R/qtl: QTL mapping in experimental crosses. *Bioinformatics.* 19:889–890.
- Calus MP, Veerkamp RF. 2011. Accuracy of multi-trait genomic selection using different methods. *Genet Sel Evol.* 43:26.
- Canaguier A, Gimpler J, Gaspero GD, Scalabrini S, Duchêne E, et al. 2017. A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genom Data.* 14:56–62.
- Chen T, Guestrin C. 2016. XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining - KDD '16.* p. 785–794. *arXiv:* 1603.02754.
- Chiquet J, Mary-Huard T, Robin S. 2017. Structured regularization for conditional Gaussian graphical models. *Stat Comput.* 27: 789–804.
- Cho S, Kim K, Kim YJ, Lee J-K, Cho YS, et al. 2010. Joint identification of multiple genetic variants via elastic-net variable selection in a genome-wide association analysis: identifying multiple variants via EN. *Ann Hum Genet.* 74:416–428.
- Condon AG, Richards RA, Rebetzke GJ, Farquhar GD. 2004. Breeding for high water-use efficiency. *J Exp Bot.* 55:2447–2460.
- CoupeL-Ledru A, Lebon C, Christophe A, Doligez A, Cabrera-Bosquet L, et al. 2014. Genetic variation in a grapevine progeny (*Vitis vinifera* L. cvs Grenache×Syrah) reveals inconsistencies between maintenance of daytime leaf water potential and response of transpiration rate under drought. *J Exp Bot.* 65:6205–6218.
- CoupeL-Ledru A, Lebon E, Christophe A, Gallo A, Gago P, et al. 2016. Reduced nighttime transpiration is a relevant breeding target for high water-use efficiency in grapevine. *Proc Natl Acad Sci USA.* 113:8963–8968.

- Cros D, Denis M, Sánchez L, Cochard B, Flori A, et al. 2015. Genomic selection prediction accuracy in a perennial crop: case study of oil palm (*Elaeis guineensis* Jacq.). *Theor Appl Genet.* 128:397–410.
- Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci.* 22:961–975.
- Dagnachew B, Meuwissen T. 2019. Accuracy of within-family multi-trait genomic selection models in a sib-based aquaculture breeding scheme. *Aquaculture.* 505:27–33.
- de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. 2013. Whole-genome regression and prediction methods applied to plant and animal breeding. *Genetics.* 193:327–345.
- Destá ZA, Ortiz R. 2014. Genomic selection: genome-wide prediction in plant improvement. *Trends Plant Sci.* 19:592–601.
- Do H, Cheon M-S, Kim SB. 2020. Graph structured sparse subset selection. *Inform Sci.* 518:71–94.
- Dodd LE, Pepe MS. 2003. Partial AUC estimation and regression. *Biometrics.* 59:614–623.
- Doligez A, Bertrand Y, Dias S, Grolier M, Ballester J-F, et al. 2010. QTLs for fertility in table grape (*Vitis vinifera* L.). *Tree Genet Genomes.* 6: 413–422.
- Doligez A, Bertrand Y, Farnos M, Grolier M, Romieu C, et al. 2013. New stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* 13:217.
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, et al. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One.* 6:e19379.
- Emanuelli F, Battilana J, Costantini L, Cunff LL, Boursiquot J-M, et al. 2010. A candidate gene association study on muscat flavor in grapevine (*Vitis vinifera* L.). *BMC Plant Biol.* 10:241.
- Ferrão LFV, Ferrão RG, Ferrão MAG, Fonseca A, Carbonetto P, et al. 2019. Accurate genomic prediction of *Coffea canephora* in multiple environments using whole-genome statistical models. *Heredity (Edinb).* 122:261–275.
- Fischer BM, Salakhutdinov I, Akkurt M, Eibach R, Edwards KJ, et al. 2004. Quantitative trait locus analysis of fungal disease resistance factors on a molecular map of grapevine. *Theor Appl Genet.* 108:501–515.
- Flutre T. 2019. rutilstimflutre: Timothee Flutre's personal R package. Zenodo. 10.5281/zenodo.3580267.
- Flutre T, Cunff L, Fodor A, Launay A, Romieu C, et al. 2020. Genome-wide association and prediction studies using a grapevine diversity panel give insights into the genetic architecture of several traits of interest. *bioRxiv.* 2020.09.10.290890.
- Fodor A, Segura V, Denis M, Neuenschwander S, Fournier-Level A, et al. 2014. Genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in Grapevine. *PLoS One.* 9:e110436.
- Fournier-Level A, Cunff LL, Gomez C, Doligez A, Ageorges A, et al. 2009. Quantitative genetic bases of anthocyanin variation in Grape (*Vitis vinifera* L. ssp. sativa) Berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics.* 183: 1127–1139.
- Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw.* 33:1–22.
- Goddard ME, Hayes BJ. 2007. Genomic selection. *J Anim Breed Genet.* 124:323–330. [_eprint:https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0388.2007.00702.x](https://onlinelibrary.wiley.com/doi/abs/10.1111/j.1439-0388.2007.00702.x)
- Gois I, Borém A, Cristofani-Yaly M, de Resende M, Azevedo C, et al. 2016. Genome wide selection in Citrus breeding. *Genet Mol Res.* 15.
- Guan Y, Stephens M. 2011. Bayesian variable selection regression for genome-wide association studies and other large-scale problems. *Ann Appl Stat.* 5:1780–1815.
- Guo G, Zhao F, Wang Y, Zhang Y, Du L, et al. 2014. Comparison of single-trait and multiple-trait genomic prediction models. *BMC Genet.* 15:30.
- Habier D, Fernando RL, Dekkers JCM. 2007. The impact of genetic relationship information on genome-assisted breeding values. *Genetics.* 177:2389–2397.
- Haley CS, Knott SA. 1992. A simple regression method for mapping quantitative trait loci in line crosses using flanking markers. *Heredity (Edinb).* 69:315–324.
- Hastie T, Qian J. 2016. *Glmnet vignette.*
- Hastie T, Tibshirani R, Friedman J. 2009. *The Elements of Statistical Learning.* Springer New York, New York, NY: Springer Series in Statistics.
- Hayashi T, Iwata H. 2013. A Bayesian method and its variational approximation for prediction of genomic breeding values in multiple traits. *BMC Bioinformatics.* 14:34.
- Heffner EL, Jannink J-L, Sorrells ME. 2011. Genomic selection accuracy using multifamily prediction models in a wheat breeding program. *Plant Genome.* 4:65–75.
- Heffner EL, Lorenz AJ, Jannink J-L, Sorrells ME. 2010. Plant breeding with genomic selection: gain per unit time and cost. *Crop Sci.* 50: 1681–1690.
- Heffner EL, Sorrells ME, Jannink J-L. 2009. Genomic selection for crop improvement. *Crop Sci.* 49:1.
- Henderson CR, Quaas RL. 1976. Multiple trait evaluation using relatives' records. *J Anim Sci.* 43:1188–1197.
- Heslot N, Yang H-P, Sorrells ME, Jannink J-L. 2012. Genomic selection in plant breeding: a comparison of models. *Crop Sci.* 52: 146–160.
- Hoerl AE, Kennard RW. 1970. Ridge regression: biased estimation for nonorthogonal problems. *Technometrics.* 12:55–67.
- Hofner B, Hothorn T. 2017. stabs: Stability Selection with Error Control. R package version R package version 0.6-3, <https://CRAN.R-project.org/package=stabs>.
- Huang Y-F, Doligez A, Fournier-Level A, Cunff LL, Bertrand Y, et al. 2012. Dissecting genetic architecture of grape proanthocyanidin composition through quantitative trait locus mapping. *BMC Plant Biol.* 12:30.
- Inouye M, Ripatti S, Kettunen J, Lyytikäinen L-P, Oksala N, et al. 2012. Novel loci for metabolic networks and multi-tissue expression studies reveal genes for atherosclerosis. *PLoS Genet.* 8: e1002907.
- Jia Y, Jannink J-L. 2012. Multiple-trait genomic selection methods increase genetic value prediction accuracy. *Genetics.* 192:1513–1522.
- Jiang C, Zeng ZB. 1995. Multiple trait analysis of genetic mapping for quantitative trait loci. *Genetics.* 140:1111–1127.
- Jiang J, Zhang Q, Ma L, Li J, Wang Z, et al. 2015. Joint prediction of multiple quantitative traits using a Bayesian multivariate ante-dependence model. *Heredity (Edinb).* 115:29–36.
- Kao C-H, Zeng Z-B, Teasdale RD. 1999. Multiple interval mapping for quantitative trait loci. *Genetics.* 152:1203–1216.
- Kemper KE, Bowman PJ, Hayes BJ, Visscher PM, Goddard ME. 2018. A multi-trait Bayesian method for mapping QTL and genomic prediction. *Genet Sel Evol.* 50:10.
- Korol AB, Ronin YI, Kirzhner VM. 1995. Interval mapping of quantitative trait loci employing correlated trait complexes. *Genetics.* 140:1137–1147.
- Kuhn M. 2008. Building predictive models in R using the caret package. *J Stat Soft.* 28:1–26.

- Kumar S, Chagné D, Bink MCAM, Volz RK, Whitworth C, et al. 2012. Genomic selection for fruit quality traits in apple (*Malus domestica* Borkh.). *PLoS One*. 7:e36674.
- Kumar S, Hilario E, Deng CH, Molloy C. 2020. Turbocharging introgression breeding of perennial fruit crops: a case study on apple. *Hortic Res*. 7:47.
- Kumar S, Kirk C, Deng CH, Shirliff A, Wiedow C, et al. 2019. Marker-trait associations and genomic predictions of interspecific pear (*Pyrus*) fruit characteristics. *Sci Rep*. 9:9072.
- Kuznetsova A, Brockhoff PB, Christensen RHB. 2017. lmerTest Package: tests in linear mixed effects models. *J Stat Soft*. 82: 1–26.
- Kwong QB, Ong AL, Teh CK, Chew FT, Tammi M, et al. 2017. Genomic selection in commercial perennial crops: applicability and improvement in oil palm (*Elaeis guineensis* Jacq). *Sci Rep*. 7: 2872.
- Lado B, Vázquez D, Quincke M, Silva P, Aguilar I, et al. 2018. Resource allocation optimization with multi-trait genomic prediction for bread wheat (*Triticum aestivum* L.) baking quality. *Theor Appl Genet*. 131:2719–2731.
- Lander ES, Botstein D. 1989. Mapping mendelian factors underlying quantitative traits using RFLP linkage maps. *Genetics*. 121: 185–199.
- Li Z, Sillanpää MJ. 2012. Overview of LASSO-related penalized regression methods for quantitative trait mapping and genomic selection. *Theor Appl Genet*. 125:419–435.
- Liu X, Hu X, Li K, Liu Z, Wu Y, et al. 2020. Genetic mapping and genomic selection for maize stalk strength. *BMC Plant Biol*. 20: 196.
- Lorenzana RE, Bernardo R. 2009. Accuracy of genotypic value predictions for marker-based selection in biparental plant populations. *Theor Appl Genet*. 120:151–161.
- Marguerit E, Brendel O, Lebon E, Leeuwen CV, Ollat N. 2012. Rootstock control of scion transpiration and its acclimation to water deficit are controlled by different genes. *New Phytol*. 194: 416–429.
- Mason L, Baxter J, Bartlett PL, Freaan MR. 1999. Boosting algorithms as gradient descent. *Adv Neural Inform Process Syst*. 12:7.
- Mayer DG, Butler DG. 1993. Statistical validation. *Ecol Model*. 68: 21–32.
- Mayor PJ, Bernardo R. 2009. Genomewide selection and marker-assisted recurrent selection in doubled Haploid versus F2 populations. *Crop Sci*. 49:1719–1725.
- McClish DK. 1989. Analyzing a portion of the ROC curve. *Med Decis Making*. 9:190–195.
- Meinshausen N, Buhlmann P. 2009. Stability selection. *J R Statist Soc B*. 72:417–473.
- Mejia N, Soto B, Guerrero M, Casanueva X, Houel C, et al. 2011. Molecular, genetic and transcriptional evidence for a role of *VvAGL11* in stenopermocarpic seedlessness in grapevine. *BMC Plant Biol*. 11:57.
- Meuwissen TH, Goddard ME. 2004. Mapping multiple QTL using linkage disequilibrium and linkage analysis information and multi-trait data. *Genet Sel Evol*. 36:261–279.
- Migicovsky Z, Sawler J, Gardner KM, Aradhya MK, Prins BH, et al. 2017. Patterns of genomic and phenomic diversity in wine and table grapes. *Hortic Res*. 4:17035.
- Muranty H, Troglio M, Sadok IB, Rifai MA, Auwerkerken A, et al. 2015. Accuracy and responses of genomic selection on key traits in apple breeding. *Hortic Res*. 2:15060.
- Nicolas SD, Péros J-P, Lacombe T, Launay A, Paslier M-CL, et al. 2016. Genetic diversity, linkage disequilibrium and power of a large grapevine (*Vitis vinifera* L.) diversity panel newly designed for association studies. *BMC Plant Biol*. 16:74.
- Obozinski G, Wainwright MJ, Jordan MI. 2011. Support union recovery in high-dimensional multivariate regression. *Ann Statist*. 39: 1–47. arXiv: 0808.0711.
- Olatoye MO, Clark LV, Wang J, Yang X, Yamada T, et al. 2019. Evaluation of genomic selection and marker-assisted selection in *Miscanthus* and energycane. *Mol Breed*. 39:171.
- Piñeiro G, Perelman S, Guerschman JP, Paruelo JM. 2008. How to evaluate models: observed vs. predicted or predicted vs. observed? *Ecol Model*. 216:316–322.
- Piepho H-P, Möhring J. 2007. Computing heritability and selection response from unbalanced plant breeding trials. *Genetics*. 177: 1881–1888.
- Price AH. 2006. Believe it or not, QTLs are accurate!. *Trends Plant Sci*. 11:213–216.
- Pszczola M, Veerkamp RF, de Haas Y, Wall E, Strabel T, et al. 2013. Effect of predictor traits on accuracy of genomic breeding values for feed intake based on a limited cow reference population. *Animal*. 7:1759–1768.
- R Core Team 2020. R: A Language and Environment for Statistical Computing. R Foundation for Statistical Computing, Vienna, Austria. Available at: <https://www.r-project.org/>.
- Rastas P. 2017. Lep-MAP3: robust linkage mapping even for low-coverage whole genome sequencing data. *Bioinformatics*. 33: 3726–3732.
- Resende MFR, Muñoz P, Resende MDV, Garrick DJ, Fernando RL, et al. 2012. Accuracy of genomic selection methods in a standard data set of loblolly pine (*Pinus taeda* L.). *Genetics*. 190: 1503–1510.
- Riedelsheimer C, Technow F, Melchinger AE. 2012. Comparison of whole-genome prediction models for traits with contrasting genetic architecture in a diversity panel of maize inbred lines. *BMC Genomics*. 13:452.
- Runcie D, Cheng H. 2019. Pitfalls and remedies for cross validation with multi-trait genomic prediction methods. *G3 (Bethesda)*. 9: 3727–3741.
- Schwander F, Eibach R, Fechter I, Hausmann L, Zyprian E, et al. 2012. Rpv10: a new locus from the Asian *Vitis* gene pool for pyramiding downy mildew resistance loci in grapevine. *Theor Appl Genet*. 124:163–176.
- Swets JA, Pickett RM, Whitehead SF, Getty DJ, Schnur JA, et al. 1979. Assessment of diagnostic technologies. *Science*. 205: 753–759.
- Thompson R, Meyer K. 1986. A review of theoretical aspects in the estimation of breeding values for multi-trait selection. *Livestock Prod Sci*. 15:299–313.
- Tibshirani R. 1996. Regression shrinkage and selection via the lasso. *J R Statist Soc B*. 58:267–288.
- Velazco JG, Jordan DR, Mace ES, Hunt CH, Malosetti M, et al. 2019. Genomic prediction of grain yield and drought-adaptation capacity in sorghum is enhanced by multi-trait analysis. *Front Plant Sci*. 10:997.
- Verzelen N. 2012. Minimax risks for sparse regressions: ultra-high dimensional phenomena. *Electron J Statist*. 6:38–90.
- Vezzulli S, Doligez A, Bellin D. 2019a. Molecular mapping of grapevine genes. In: D Cantu, MA Walker, editors. *The Grape Genome*. Cham: Compendium of Plant Genomes Springer International Publishing. p. 103–136.
- Vezzulli S, Zulini L, Stefanini M. 2019b. Genetics-assisted breeding for downy/powdery mildew and phylloxera resistance at fem. *BIO Web Conf*. 12:01020.

- Viana AP, Resende M. D V D, Riaz S, Walker MA. 2016a. Genome selection in fruit breeding: application to table grapes. *Sci Agric (Piracicaba, Braz)*. 73:142–149.
- Viana JMS, e Silva FF, Mundim GB, Azevedo CF, Jan HU. 2016b. Efficiency of low heritability QTL mapping under high SNP density. *Euphytica*. 213:13.
- Waldmann P, Mészáros G, Gredler B, Fuerst C, Sölkner J. 2013. Evaluation of the lasso and the elastic net in genome-wide association studies. *Front Genet*. 4:270.
- Wang F, Mukherjee S, Richardson S, Hill SM. 2020. High-dimensional regression in practice: an empirical study of finite-sample prediction, variable selection and ranking. *Stat Comput*. 30:697–719.
- Wang Y, Mette MF, Miedaner T, Gottwald M, Wilde P, et al. 2014. The accuracy of prediction of genomic selection in elite hybrid rye populations surpasses the accuracy of marker-assisted selection and is equally augmented by multiple field evaluation locations and test years. *BMC Genomics*. 15:556.
- Welter LJ, Göktürk-Baydar N, Akkurt M, Maul E, Eibach R, et al. 2007. Genetic mapping and localization of quantitative trait loci affecting fungal disease resistance and leaf morphology in grapevine (*Vitis vinifera* L). *Mol Breed*. 20:359–374.
- Wickham H. 2016. *ggplot2: Elegant Graphics for Data Analysis*. New York, NY: Springer-Verlag.
- Yi H, Breheny P, Imam N, Liu Y, Hoeschele I. 2015. Penalized multi-marker vs. single-marker regression methods for genome-wide association studies of quantitative traits. *Genet Sec Invest*. 199: 205–222.
- Zou H, Hastie T. 2005. Regularization and variable selection via the elastic net. *J R Statist Soc B*. 67:301–320.

Communicating editor: J-L. Jannink