



Enhancing grapevine breeding efficiency through genomic prediction and selection index

Charlotte Brault, Vincent Segura, Maryline Roques, Pauline Lamblin, Virginie Bouckennooghe, Nathalie Pouzalgues, Constance Cuntty, Matthieu Breil, Marina Frouin, Léa Garcin, et al.

► To cite this version:

Charlotte Brault, Vincent Segura, Maryline Roques, Pauline Lamblin, Virginie Bouckennooghe, et al.. Enhancing grapevine breeding efficiency through genomic prediction and selection index. G3, 2024, 14 (4), pp.jkae038. 10.1093/g3journal/jkae038 . hal-04563189

HAL Id: hal-04563189

<https://hal.inrae.fr/hal-04563189>

Submitted on 29 Apr 2024


HAL is a multi-disciplinary open access archive for the deposit and dissemination of scientific research documents, whether they are published or not. The documents may come from teaching and research institutions in France or abroad, or from public or private research centers.

L'archive ouverte pluridisciplinaire **HAL**, est destinée au dépôt et à la diffusion de documents scientifiques de niveau recherche, publiés ou non, émanant des établissements d'enseignement et de recherche français ou étrangers, des laboratoires publics ou privés.



Distributed under a Creative Commons Attribution - NonCommercial 4.0 International License

Enhancing grapevine breeding efficiency through genomic prediction and selection index

Charlotte Brault ^{1,2,*} Vincent Segura,^{1,3} Maryline Roques,^{1,2} Pauline Lamblin,² Virginie Bouckennooghe,^{1,2} Nathalie Pouzalgues,⁴ Constance Cuntz,^{2,4} Matthieu Breil,^{1,2} Marina Frouin,⁵ Léa Garcin,^{2,5} Louise Camps,⁵ Marie-Agnès Ducasse,² Charles Romieu,^{1,3} Gilles Masson,^{2,4} Sébastien Julliard,⁵ Timothée Flutre,⁶ Loïc Le Cunff^{1,2,*}

¹UMT Geno-Vigne®, IFV, INRAE, Institut Agro Montpellier, Montpellier 34398, France

²Institut Français de la vigne et du vin, Pôle National Matériel Végétal, Le Grau du Roi 30240, France

³UMR AGAP Institut, Univ Montpellier, CIRAD, INRAE, Institut Agro Montpellier, Montpellier 34398, France

⁴Centre du Rosé, Viduban 83550, France

⁵Conservatoire du Vignoble Charentais, Institut de Formation de Richemont, Cherves-Richemont 16370, France

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

*Corresponding author: Email: loic.lecunff@vignevin.com; *Corresponding author: Email: charlotte.brault@live.com

Grapevine (*Vitis vinifera*) breeding reaches a critical point. New cultivars are released every year with resistance to powdery and downy mildews. However, the traditional process remains time-consuming, taking 20–25 years, and demands the evaluation of new traits to enhance grapevine adaptation to climate change. Until now, the selection process has relied on phenotypic data and a limited number of molecular markers for simple genetic traits such as resistance to pathogens, without a clearly defined ideotype, and was carried out on a large scale. To accelerate the breeding process and address these challenges, we investigated the use of genomic prediction, a methodology using molecular markers to predict genotypic values. In our study, we focused on 2 existing grapevine breeding programs: Rosé wine and Cognac production. In these programs, several families were created through crosses of emblematic and interspecific resistant varieties to powdery and downy mildews. Thirty traits were evaluated for each program, using 2 genomic prediction methods: Genomic Best Linear Unbiased Predictor and Least Absolute Shrinkage Selection Operator. The results revealed substantial variability in predictive abilities across traits, ranging from 0 to 0.9. These discrepancies could be attributed to factors such as trait heritability and trait characteristics. Moreover, we explored the potential of across-population genomic prediction by leveraging other grapevine populations as training sets. Integrating genomic prediction allowed us to identify superior individuals for each program, using multivariate selection index method. The ideotype for each breeding program was defined collaboratively with representatives from the wine-growing sector.

Keywords: genomic prediction; grapevine; plant breeding; selection index; ideotype; Cognac; Rosé; genomic selection; Genomic Prediction; GenPred; Shared Data Resource

Introduction

Plant breeding has been a key lever to adapt varieties to human use and the environment. The genetic gain obtained after 1 cycle of a breeding program is given through the breeder's equation (Lush 1937). It depends on the additive genetic variance of the population, the accuracy and intensity of selection, and the cycle length. In grapevine, this cycle length is about 20–25 years, when accounting for phenotyping new varieties (Töpfer and Trapp 2022). Thus, the long breeding cycle for grapevine breeding is a major constraint to genetic gain. Because of its perennial nature, grapevine needs to be adapted to challenging conditions, in an increasingly variable environment, due to climate change (Santos et al. 2020).

In the past years, grapevine breeding in Europe has been focused on disease resistance to powdery and downy mildews (Eibach et al. 2007; Schneider et al. 2019; Töpfer and Trapp 2022). The French INRAE-ResDur program generated a dozen of varieties, all with at least 2 major resistance genes for each disease.

The whole selection process lasted around 15–20 years (Reynolds 2015; Schneider et al. 2019). Thus, there is a critical need for accelerating this selection process while accounting for other traits related to climate change. Marker-assisted selection (MAS) was used in the INRAE-ResDur program for early screening of seedlings with major resistance genes for powdery and downy mildews. However, most traits involved in adaptation are under a complex genetic determinism with possibly thousands of genes involved, including polygenic resistance, wine quality, and yield (Alonso-Blanco and Méndez-Vigo 2014; Flutre et al. 2022). For such traits, QTL detection results in many small effects often overestimated and that are not transferable through MAS to breeding (Beavis et al. 1994; Xu 2003; Meuwissen et al. 2016; Crossa et al. 2017).

Genomic selection (GS) has been proposed to avoid these limitations, thanks to the availability of genome-wide markers (Bernardo 1994; Meuwissen et al. 2001). In GS, all markers are analyzed together and their associated effects on the phenotypes are jointly estimated in a training set (TS) population. Then, these

effects are applied in a validation set (VS) population, on which only genotypic data are available (Heffner et al. 2009). GS has been widely applied to animal and plant breeding, with some few examples of applications in grapevine (Fodor et al. 2014; Viana et al. 2016; Migicovsky et al. 2017; Brault et al. 2021, Brault, Segura, et al. 2022; Flutre et al. 2022). Notably, GS has only been applied in a research context, with varieties not intended for breeding. GS can save time in breeding programs, but it offers other benefits too (R2D2 Consortium et al. 2021). Indeed, using GS allows testing of more crosses and offspring because no phenotyping is needed. This increases the selection intensity, as more genotypes are tested, increasing the selection gain according to the breeder's equation. Concerning the selection accuracy, the impact of GS is balanced. On the one hand, GS implies concentrating phenotyping on the training population, with possibly more replications that can increase the heritability and accuracy of the model. On the other hand, using a GS model trained in a population genetically far from the selection population would reduce the predictive ability (PA) (Brault, Segura, et al. 2022). One challenge of GS is then to find a trade-off between the advantages and drawbacks of GS in terms of prediction accuracy.

Once predicted or observed genotypic values are acquired, the breeder needs to select the best individuals in the population, by taking into consideration several traits and making compromises. This can be streamlined with a linear multitrait selection index. The most famous selection index is the Smith–Hazel index (Smith 1936). Since then, other algorithms have been developed to account for the multicollinearity between the traits (de Carvalho Rocha et al. 2018; Olivoto and Nardino 2021). In grapevine, the ideotype (i.e. the criteria to combine all traits to get the best performing variety in each environment) is complex, because the wine is a transformed product and its quality relies on many variables (Reynolds 2015; Töpfer and Trapp 2022). Such an ideotype is likely to vary across wine regions. Specifically, the grapevine ideotype will include traits for which the genetic value must be maximized or minimized (directional selection) and traits for which an optimum value would be sought (stabilizing selection). Moreover, quality traits such as acids, sugars, anthocyanins, tannins, and volatile compounds interact with yield-related variables (Reynolds 2015).

This article describes and proposes an application of GS to 2 breeding programs of grapevine varieties. These 2 breeding programs were compared, with a similar design of experiments but various traits and ideotypes. First, we fitted a mixed linear model for each experiment to extract genotypic values, and then, we applied genomic prediction (GP) within the TS to estimate PA. Finally, we used multitrait selection index to select the most promising individuals from predicted genotypic values.

Material and methods

Design of experiment

Two breeding programs (hereafter referred to as population) were compared: the Martell breeding program, funded by Martell company which produces Cognac and conducted by the conservatory of the Charente vineyards, INRAE, and IFV in France; and the EDGARR breeding program, conducted by the Center for Rosé, INRAE, and IFV in France for producing Rosé wine. Both programs included biparental families, hereafter referred to as crosses between varieties emblematic of the region and varieties with polygenic resistance to powdery and downy mildews (interspecific hybrids). Resistant varieties were obtained by crossing *Vitis vinifera* varieties with other *Vitis* wild species bearing diverse

resistance genes. Then, several cycles of backcross with *V. vinifera* varieties were done to retrieve the quality of *V. vinifera* (Töpfer et al. 2010). The resulting crosses are called interspecific. In both programs, after MAS, unselected individuals from the crosses were planted in a pot to constitute the TS, for genotyping and phenotyping; while selected individuals (i.e. with all resistance genes based on MAS) were only genotyped and constitute the VS, a few families were only in the VS (Supplementary Fig. 1, Fig. 1). This specific design was applied to speed up the breeding program.

The Martell program included 4 famous grape varieties (Monbadon, Montils, Rayon d'Or, and Vidal 36) crossed with 5 different resistant varieties (E03-PL4, C03-PL5, 50#90, 50#86, 50#83), and the EDGARR experiment included 2 famous grape varieties (Cinsaut and Vermentino) crossed with 6 resistant varieties (mainly 3421-F02-PL5 and 3408-F10-PL2 and COL48#59, COL48#46, COL48#29, COL13#75). The detailed composition of the TSs and VSs is in Supplementary Fig. 1. The genetic relatedness between the individuals of the TS and the VS could be full-sibs, half-sibs, or no genetic relationship. A major difference between these programs was the number of genotypes. In the Martell program, there were 347 and 277 individuals in the TS and VS, respectively. In the EDGARR program, there were 193 and 132 individuals in the TS and VS, respectively.

Genomic data analysis

The same genotyping approach was used in both programs. Genotyping was done using the genotyping-by-sequencing technology, using the ApeKI restriction enzyme (Elshire et al. 2011). Paired-end sequencing (2 × 150 bp) was performed on 4 lanes of an Illumina HiSeq2000 for each program (at the MGX platform in Montpellier, France). Processing consisted in read checking with FastQC version 0.1.2 (Andrews 2016), demultiplexing with a custom script, cleaning and trimming with CutAdapt version 1.8.1 (Martin 2011), alignment on the PN40024 12Xv4.2 reference sequence (Velt et al. 2023) with BWA-MEM version 0.7.12-r1039 (Li 2013) and realignment with GATK version 3.7 (DePristo et al. 2011), followed by variant and SNP calling with GATK HaplotypeCaller, and a final filtering step, notably to discard SNP genotypes with <10 reads or quality below 20. For EDGARR and Martell programs, SNP markers with more than 10% missing data and with <20 reads were discarded, producing 27,271 and 10,602 remaining SNPs, respectively. Genotypes with more than 50% missing data were also discarded. The remaining SNPs were imputed using Beagle software version 5.4 (Browning et al. 2018). Markers with a minor allele frequency lower than 1% were removed, giving a final table for EDGARR of 19,228 SNP markers for 326 individuals and for Martell of 10,380 SNPs for 624 individuals.

Then, for each cross, outlying individuals were detected using the Mahalanobis distance (Mahalanobis 1936), with a P-value of 1%.

Phenotypic data analysis

EDGARR

The individuals in the training population (e.g. 193 genotypes) were planted in pots, without rootstock. The vinestocks were managed to fast fruiting to accelerate the production of grapes. The experimental trial was located at the Espiguette domain, in Grau-du-Roi, in the South of France (43°29'48.5"N 4°08'13.2"E). There was 1 repetition per genotype, except for repeated controls Cinsaut, Vermentino, Grenache, and Syrah, with 5 or 6 repetitions; the design is close to a p-rep design with a large proportion of

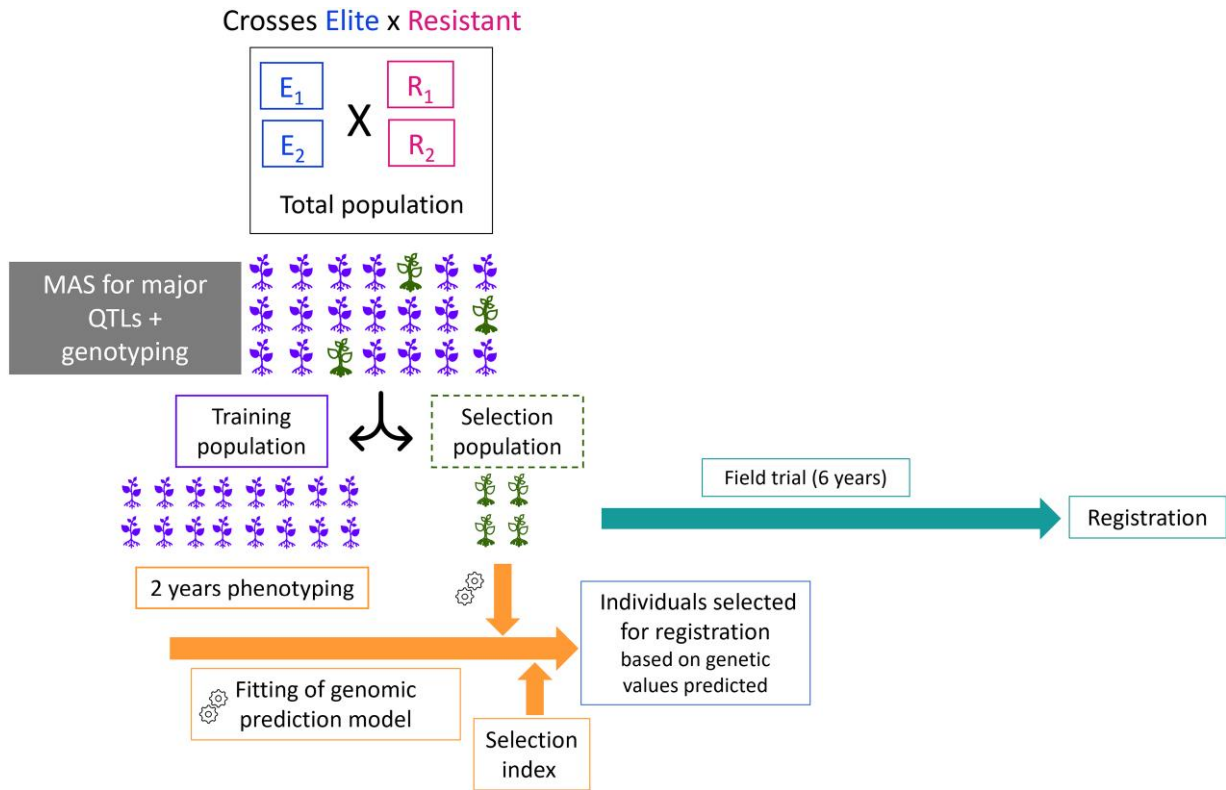


Fig. 1. Design of experiment for EDGARR and Martell breeding programs. Green (empty) seedlings carry 2 resistance genes for both powdery and downy mildews; purple seedling are missing resistance genes.

individuals unreplicated. This design was used to save time, because replication requires clonal propagation with cuttings, which takes a lot of time from seedling stage. Genotypes were mostly ordered by crosses in the field layout for practical reasons, with controls randomly repeated.

In this population, 30 traits were phenotyped for 2 years (2018 and 2019), and 5 additional traits were phenotyped for 1 year. Traits were divided into 5 categories, namely acids with cis- and trans-coutaric acids, caftaric, ascorbic, hydroxycinnamic, malic, shikimic, and tartaric acids, pH, and total acidity; color traits with blue, yellow, and red absorbance, lightness, yellow and red indices, color intensity, and tint and polymeric pigments at 420 and 520 nm; sugar traits with glucose and fructose; polyphenol traits with total polyphenol index and anthocyanin concentration; and finally agronomic and technologic traits with berry weight, glutathione, number of clusters, and harvest date. A full description of these traits and summary statistics are [Supplementary Tables 1 and 2](#). Clusters were sampled when the sugar content reached 22° brix (gram of saccharose/100 g). Some traits were measured with 2 nonredundant units: in concentration (g/L) and in amount in berries (mg/g of berries).

For the extraction of genotypic values, we first applied a full mixed model for each trait phenotyped for 2 years:

$$y_{ijkl} = \mu + G_i + C_j + x + y + Y_k + \epsilon_{ijkl}, \quad (1)$$

with y_{ijkl} the phenotypic observation for a given genotype i , cross j , year k , and repetition l (for controls), μ the intercept, G_i the random effect of the genotype j (nested in cross i), C_j the random effect of the cross (number of genotypes per cross is available in [Supplementary Fig. 1](#)), x and y the random effects for coordinates

(i.e. row and column with 4 and 95 levels, respectively) of the pots in the trial, Y_k the fixed effect of the year (2 levels), and ϵ_{ijkl} the residuals. We assume that G_i , C_j , x , y , and ϵ_{ijkl} are random, with independent distribution and mean of 0, and have the following variances σ_G^2 , σ_C^2 , σ_x^2 , σ_y^2 , and σ_ϵ^2 . The total genetic variance is $\sigma_g^2 = \sigma_G^2 + \sigma_C^2$; a homogeneous genotype and cross variance was assumed. This full model was fitted with maximum likelihood, random effects were selected by a likelihood ratio test, and fixed effects were selected based on Fisher tests, using the lmerTest R package ([Bates et al. 2014](#)). The homogeneity of cross variances was checked visually. Variance components were estimated with restricted maximum likelihood on the selected model. The broad-sense heritability (H^2), was computed as:

$$H^2 = \frac{\sigma_G^2 + \sigma_C^2}{\sigma_G^2 + \sigma_C^2 + \sigma_\epsilon^2} \quad (2)$$

with σ_G^2 , σ_C^2 , and σ_ϵ^2 variances associated with genotype, cross, and residuals. Fitting information for all traits is available in [Supplementary Table 3](#). If the genotype effect was not selected in the model, we refitted the same model with model selection only for fixed effects.

Best Linear Unbiased Predictors (BLUPs) were computed as the sum of the genotypic and cross effects (when cross effect was selected). We deregressed the BLUPs with the following formula:

$$\text{drgBLUP} = \frac{\text{BLUP}}{1 - (\text{PEV}/\sigma_g^2)} \quad (3)$$

([Garrick et al. 2009](#); [Andrade et al. 2019](#)), with PEV the prediction error variance, i.e. the error associated with each BLUP value (for

genotype and cross effects). This was estimated by the “postVar” parameter in ranef function from the lme4 R package. The denominator of Equation 3, i.e. the reliability of BLUP, was variable among genotypes because some of them had phenotypic data only for 1 year depending on the trait, and controls had more repetitions. For traits measured for 1 year, raw phenotypic data were used (averaged phenotypic data per genotype was computed for checks).

Martell

For the Martell program, individuals were also planted in pots for the training population, without rootstocks, in Cognac region (45°44'22.9"N 0°21'58.2" E). The TS included 358 genotypes; among them, 349 came from progenies, and 9 were grafted field controls (repeated 5 times). The phenotyping was done in 2021 and 2022 on potted plants for the training population. We studied 30 traits, which can be classified into 6 categories: vigor, disease, phenology, agronomic, technologic, and vinification. A full description of these traits and summary statistics are in Supplementary Tables 1 and 2. Traits related to harvest were sampled at around 10 alcohol content for the referent genotype (Ugni blanc).

The mixed model equation for phenotypic data analysis included effects described in Equation (1) and some other effects: $y_{ijkl} = \mu + G_i + C_j + x + y + Y_k + Rpv3_i + Ren3_i + Run1_Rpv1_i + M_{ijkl} + \epsilon_{ijkl}$, with the supplementary random effects for resistance genes Rpv3, Ren3, Run1_Rpv1, and M indicating the presence of available vine spur (if 1 spur and 1 cane were present, the pruning was simple guyot). Martell field design contained 16 rows and 26 columns (for x and y effects). We used the same Equation (2) for computing the heritability for Martell population.

GP

The same pipeline of analysis was applied to both programs. First, Genomic Prediction was applied to the training population for all traits available with K-fold cross-validation, repeated $R = 10$ times, with $K = 5$. We implemented 2 GP methods: GBLUP with rrBLUP R package (Endelman 2011) and the LASSO (Tibshirani 1996), with glmnet R package (Friedman et al. 2010). GBLUP is more adapted to a complex genetic architecture (many QTLs), while LASSO is more adapted to a simpler genetic architecture. PA was estimated as Pearson's correlation between observed and the predicted genotypic values. PA values were averaged across folds and cross-validation repetitions, and standard errors were calculated.

The best method among the 2 was chosen for each population and trait and used to predict the genotypic values for the VS. The model was refit on the whole TS (without cross-validation) for predicting genotypic values for the VS. These values were deregressed a second time using Equation (3), with genetic variance and PEV estimated using mixed.solve function from rrBLUP R package (Endelman 2011). For the LASSO, the deregressed values were obtained by fitting the ordinary least square estimator for all selected markers in the TS.

For the EDGARR experiment, we predicted the berry color (red or white) using a logistic generalized linear model (GLM), adapted to binomial data with the LASSO method, using the glmnet R package, with as options family='binomial' and alpha=1 (Friedman et al. 2010).

Selection index

The selection index was designed by representatives of the wine growers for each of the 2 studied wine regions. It included traits for which the value needs to be maximized or minimized and traits for which an optimal value is required. The first selection criterion was the presence of the resistance genes for powdery and downy mildews and the flower sex, handled with MAS.

The resulting multivariate selection index was computed using the multitrait genotype-ideotype distance index (MGIDI) method, described in Olivoto and Nardino (2021). Briefly, it rescales the phenotype on a 0–100 scale, in which 100 represents the maximum or the minimum value, depending on the direction of the selection. Then, it performs a factor analysis, to summarize the multitrait phenotypes and to avoid collinearity. Finally, the MGIDI is given by the sum of the distance between the actual phenotype and the ideotype for each factor. When an optimal value was sought by professionals, we computed the difference between the optimal value and the phenotype.

The selection index was applied for both programs, on predicted and deregressed genotypic values for the VS individuals. The output of the MGIDI method included a strength and weakness view of selected individuals, with the contribution of each factor to the distance to the ideotype, and the rank of individuals, ordered by increasing MGIDI value.

Other phenotypic and genomic data

We used genomic and phenotypic data from 2 other grapevine populations. A half-diallel population composed of 628 individuals from 10 biparental crosses where 5 parents were involved (Tello et al. 2019), phenotyped between 2013 and 2017. The second population is a diversity panel population of 277 genotypes, chosen to represent the genetic diversity of *V. vinifera* (Nicolas et al. 2016) and phenotyped between 2011 and 2012. Phenotypic and genomic data from these populations were already analyzed for GP and QTL detection in previous studies (Brault, Lazerges, et al. 2022, Brault, Segura, et al. 2022; Flutre et al. 2022).

There were 6 and 5 common traits with EDGARR and Martell programs, respectively. For genomic data, we performed a Basic Local Alignment Research Tool (BLAST) analysis on flanking sequences to find out the marker positions corresponding to the last version (PN40024.v4) of the *V. vinifera* reference genome (Velt et al. 2023). Then, we kept the common markers between each population and the target one (Supplementary Table 4). We fitted a GP model using GBLUP and LASSO for half-diallel, diversity panel, or both populations and kept the best method to predict genotypic values of EDGARR and Martell populations. We measured the PA and compared it with the values from within-population GP (i.e. when cross-validation was applied within EDGARR or within Martell genotypes). Additive genetic relatedness between each pair of genotype in TS and VS was assessed using rrBLUP R package version 4.6.2 (Endelman 2011).

Results

Genetic structure

For the EDGARR program, 325 individuals were genotyped for 19,228 SNP markers after filters. For the Martell program, there were 624 individuals genotyped for 10,380 SNPs. A principal component analysis (PCA) was conducted to explore the genetic structure of the populations. We found that families were well separated, located between their parents (Fig. 2). Individuals in TS and VS displayed a clear overlap, except for some families only in the VS (Fig. 2, Supplementary Fig. 1). The PCA analysis showed some outlier individuals, spotted with the Mahalanobis distance. For the EDGARR population, we excluded 3 individuals, all from Cinsaut × 3421-F02-PL5 cross; for the Martell population, we excluded 4 individuals from 4 crosses.

Overall, for both populations, the relative position of biparental families seems to be driven by the resistant parents (3408-F10-PL2, 3421-F02-PL5, 50#86, C03-PL5, E03-PL4), as crosses from a common resistant parent are grouped together. This is likely

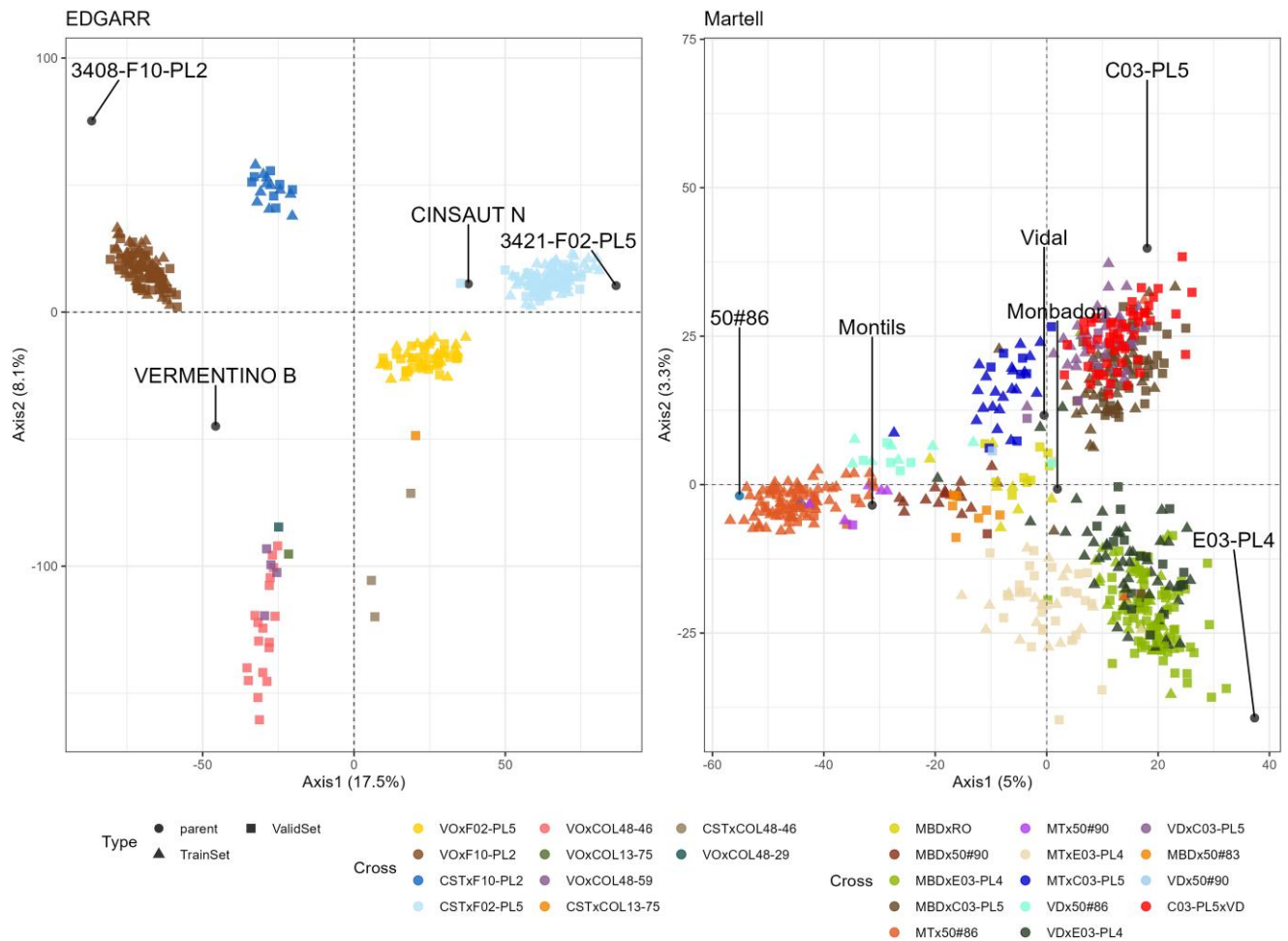


Fig. 2. Principal Component Analysis of genetic markers for EDGARR (left panel) and Martell (right panel) populations. Parents are labeled. The point shape corresponds to the type of individual: triangle: Training Set; square: Validation Set; points: parents of crosses. Cross names were abbreviated as follows: Vermentino (VO), Cinsaut (CST), Monbadon (MBD), Rayon d'Or (RO), Montils (MT), and Vidal (VD).

because those resistant parents are likely more genetically distant between themselves compared with *V. vinifera* varieties.

Phenotypic structure

In phenotypic data analysis and for both populations, we included nongenetic factors, despite a small number of repetitions, and those effects were significant for some traits (Supplementary Table 3). Broad-sense heritability values displayed a wide range across all traits (Supplementary Table 3). They ranged from 0 to 0.76 (average of 0.39) and from 0.002 to 0.99 (average of 0.43) for EDGARR and Martell populations, respectively. From the BLUPs of genotypic values, we applied a deregression to retrieve the original scale of the data in terms of mean and variance. We checked visually the quality of deregression. The Pearson correlation between raw averaged phenotypic data and deregressed BLUPs for each trait was between 0.73 and 0.98 for the EDGARR population and between 0.60 and 0.99 for the Martell population. Besides that, we measured the matrix of genotypic correlations (for deregressed BLUPs) between the traits. We found that for EDGARR, traits related to color were more correlated to each other, with correlation values from 0.74 to 0.99 (those types of traits were not present in Martell data set). Overall, for the other traits, genotypic correlations were mostly low (data not shown).

The PCA analysis showed a mild phenotypic structure (Supplementary Fig. 2). For EDGARR, the structure was driven by

the resistant parents (3421-F02-PL5 and 3408-F10-PL2) and by traits related to color, while for Martell, the crosses were more separated from each other, and the differentiation was driven by acid and yield traits.

GP results

We used a GLM with the LASSO method to predict categorical color for EDGARR population. The accuracies ranged between 0.943 and 0.963, with an average of 0.952 in cross-validation. For the other traits, predictive abilities were comparable for both populations and covered a wide range of values between 0.04 and 0.87 (Fig. 3). To avoid the effect of the genetic architecture on the PA, we chose the best method between GBLUP and LASSO. Overall, GBLUP provided a better PA than LASSO for both populations, with an average of 0.41 and 0.34 for EDGARR and 0.44 and 0.39 for Martell, for GBLUP and LASSO, respectively. For EDGARR, GBLUP yielded a higher PA than LASSO for 26 traits out of 35 and 28 out of 30 for Martell. For both populations, PA was correlated with broad-sense heritability values across traits, with a correlation value of 0.60 for EDGARR and 0.42 for Martell. The different trait categories were quite evenly represented across the range of PA for both populations (Fig. 3). However, traits for which the cross effect was not kept in the mixed model (Equation 1) displayed a lower PA with an average difference of 0.55 and 0.37 for EDGARR and Martell populations, respectively (Supplementary Fig. 3). Thus, the population structure of the

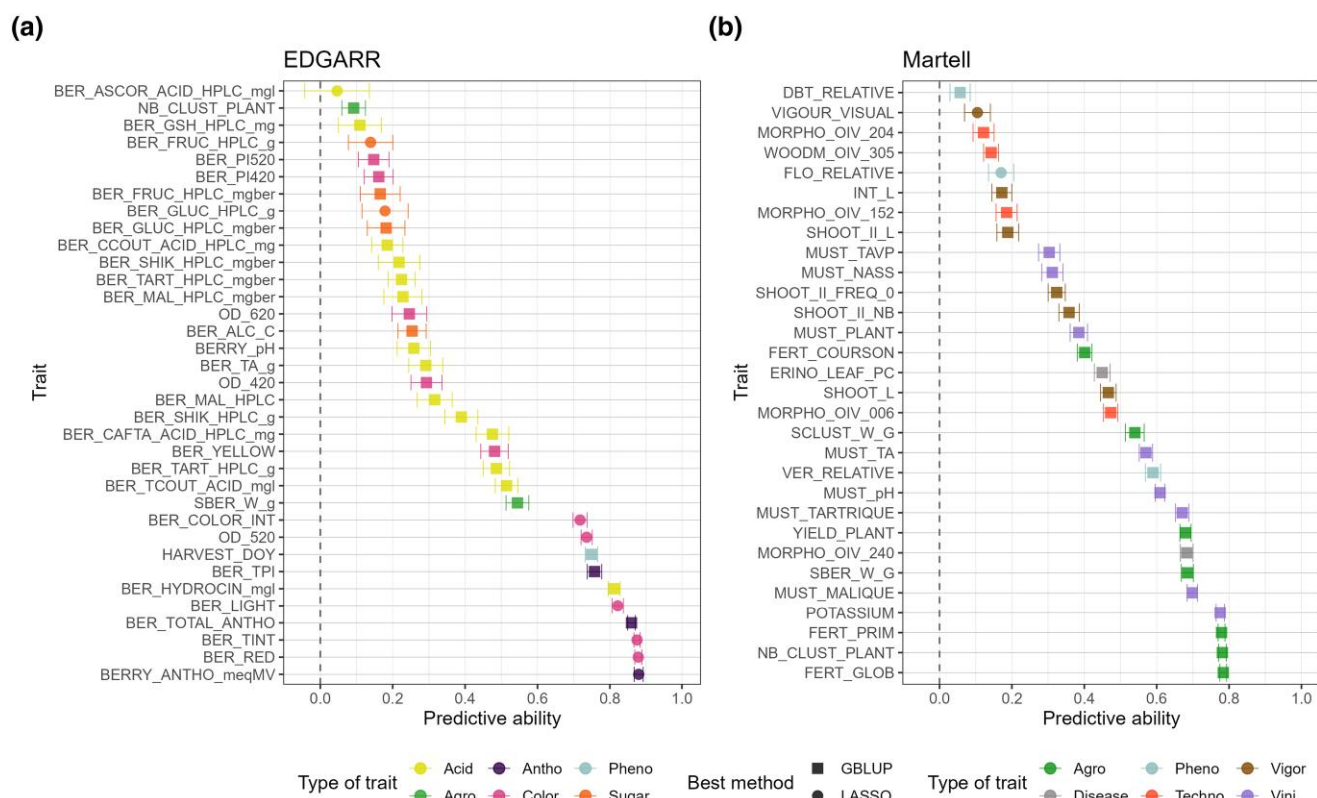


Fig. 3. PA for all traits for EDGARR (a) and Martell (b) populations. Error bars correspond to standard errors calculated across cross-validation repetitions. For each trait, the best method among GBLUP and LASSO was selected.

BLUPs has a strong impact on accuracy. We found that the 4 traits measured on a semi-quantitative scale for Martell populations had a slightly lower PA (difference of 0.26, a *P*-value of 0.044 using a Wilcoxon test). For EDGARR data, we could not fit a mixed model for 6 traits, because they were phenotyped in a single year. We found an average PA of 0.17 for these traits, up to 0.51 for trans-coutaric acid and the GBLUP method.

The predicted genotypic values were deregressed a second time in order to retrieve the initial mean and variance for each trait for applying our selection index. We computed the correlation between the genotypic values (deregressed BLUPs) and the predicted values (obtained by GP) and visually checked that the scales were comparable. The correlations ranged from 0.42 to 1 (average of 0.82) for the EDGARR population and from 0.24 (for vigor trait) to 1 (average of 0.87) for the Martell population.

Selection index

For both programs, the selection index was established by the professional committee in charge of local grapevine breeding. The first criterion was the presence of 2 resistance genes both to powdery and downy mildews. Then, a specific index was determined, based on the traits available.

EDGARR selection index

For EDGARR population, the *Center of Rosé* established a selection index to get varieties with more acidity, less color, higher productivity, and adaptation to climate change. Finally, the corresponding ideotype included 11 traits, 5 traits to be optimized (must tartaric, malic, total acidity, pH, alcohol content), 2 traits to be minimized (color intensity and total polyphenol index), and 4 traits to be

maximized (berry tint, number of clusters, berry weight, and harvest date) (Supplementary Table 5). We used PA values as weights associated with each trait. The MGIDI algorithm selected 3 factors, represented by 4 (tartaric acid, berry lightness, berry color, color intensity, harvest date), 3 (malic acid, total acidity, pH), and 4 traits (alcohol content, total polyphenol index, number of cluster, and berry weight), respectively (Supplementary Table 6). Distributions of predicted genotypic values and the position of some parents and selected genotypes are displayed in Supplementary Fig. 4. Vermentino was a parent of 12 out of 15 selected genotypes, and 8 individuals from the same cross Vermentino × F10-PL2 were selected (Supplementary Table 7). Surprisingly, the resistant genotype F02-PL5 was not selected as a parent of the first 15 genotypes. From the PCA and distribution analysis (Fig. 4, Supplementary Fig. 4), it is clear that selected individuals are phenotypically close to each other. The predicted berry color was white for 4 genotypes, and the genotype with the lowest MGIDI was predicted white (Supplementary Table 7). Factors 1 and 2 contributed the most to the MGIDI score for the selected genotypes, which means that they performed quite similarly for factor 3. Some genotypes performed better for some factors, such as P869-F04 for factor 1 or P596-A09 for factor 2, while others had a more balanced performance across factors, such as P249-F10 (Supplementary Fig. 5, Supplementary Table 6).

Martell selection index

The ideotype for Martell included 12 traits, 7 traits to be maximized (global and primary fertility, yield, tartaric acid, total must acidity, cluster weight, and berry weight), 1 to be minimized (must pH), and 2 with an optimum value (must malic acid and ease of detachment of pedicel, OIV 240) (Supplementary

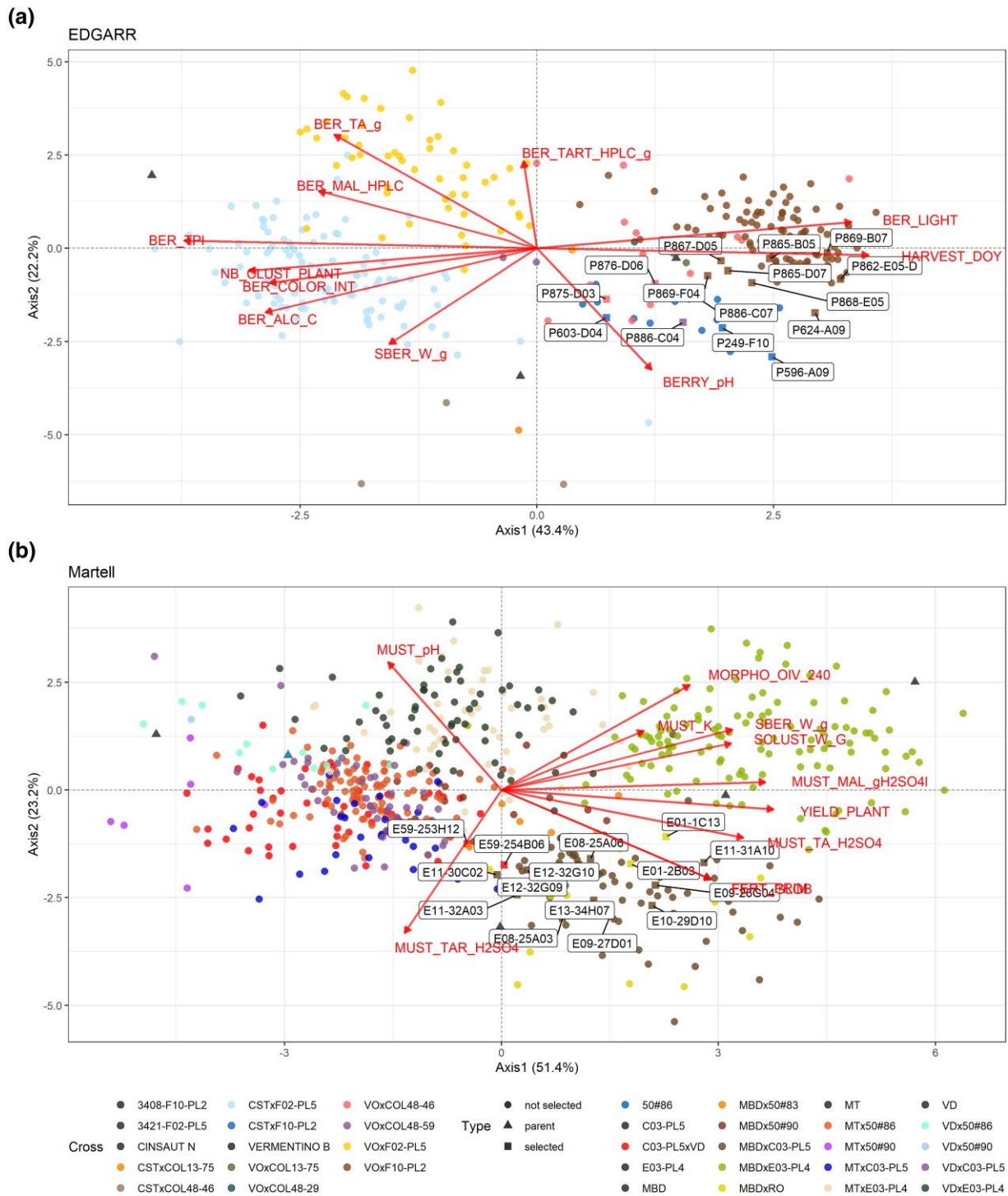


Fig. 4. PCA of the genotypic values for the selection candidates for the traits in the selection index for the first 2 principal components. Variables are displayed in red arrows, and genotypes are colored according to their cross. Selected individuals are labeled. a) EDGARR population and b) Martell population.

Table 5). We excluded beforehand traits with a PA value lower than 0.5. The MGIDI algorithm selected 3 factors, represented by 5 (total acidity, pH, yield, and primary and global fertility), 4 (cluster weight, tartaric acid, berry weight, and malic acid), and 2 traits (ease of detachment of pedicel and potassium), respectively (Supplementary Table 5). Among the selected traits, some of them displayed high genetic correlations (positive or negative). The average of the 15 genotypes selected followed the expected trend (increase or decrease compared with the average of the

population), for all the traits, except for single berry, cluster weights, and malic acid (Supplementary Table 6). Distributions of predicted genotypic values and the position of some parents and selected genotypes are displayed in Supplementary Fig. 4. For 13 out of 15 genotypes selected, Monbadon and C03-PL5 were 1 of the 2 parents (Supplementary Table 7). As for EDGARR, factor 3 contributed less to the MGIDI score, and genotypes displayed various strengths or weaknesses for the factors. In particular, the superior performance of E12-32G10 (ranked 1st) was due to factors

1 and 3, and E10-29D10 (ranked 8th) was only due to factors 2 and 3 (Supplementary Fig. 5, Supplementary Table 6).

Across-population GP

We studied the ability of nonrelated populations (the half-diallel and the diversity panel populations) to predict genotypic values for EDGARR and Martell population. For that, we compared the PA of each population from Fig. 3 (referred to as within-population), with PAs for the half-diallel or the diversity panel as a training population (referred to as across-population). For EDGARR and Martell, within-population GP was better for 4 traits out of 6 and for 5 traits out of 5, respectively. PA values for across-population GP were variable, mostly depending on the trait, on the validation population, and to a lesser extent on the training population (Fig. 5). Overall, across-population PA values were much higher in EDGARR than in the Martell program. For EDGARR and 2 traits (shikimic acid concentration and number of clusters), using data from the diversity panel and the half-diallel led to a higher PA than using data from the same population (EDGARR).

For EDGARR, using both data from the diversity panel and the half-diallel led to higher PA, except for the harvest date, for which a strong decrease was observed. For Martell, the diversity panel was the best training population consistently for all traits. When looking closely at predicted values using the different training population we studied (Supplementary Fig. 6), we noticed that genetic structure (among EDGARR or Martell crosses) was often better captured for EDGARR than for Martell population in across-population. This led to an improvement in PA, notably for shikimic acid, tartaric acid, and single berry weight.

We used for each trait and TS the best method among GBLUP and LASSO. The results show that GBLUP was the only method selected for within-population, while LASSO was the best method for some traits and training populations (Supplementary Fig. 6).

The difference in across-population PAs between the EDGARR and Martell programs is consistent with a higher genetic relatedness between the training populations tested (half-diallel and diversity panel) and EDGARR (Supplementary Fig. 7). Interestingly, the additive genetic relatedness between half-diallel crosses or diversity panel subpopulations and EDGARR crosses was more contrasted than with Martell crosses (Supplementary Fig. 7).

Discussion

Our study comprised the analysis of 30 traits for 2 grapevine populations. Some of the individuals were only genotyped, which allowed us to perform GPs. We first tested the ability of GP models to accurately predict the genotypic values in a within-population scenario. Then, we proposed a selection index and selected the most relevant individuals according to it. The ideotype was built in partnership with professional wine growers and was specific to each of the 2 wine regions studied. To our knowledge, this is the first time a precise ideotype is described for grapevine. Finally, some of the phenotyped traits were also available for other grapevine populations. We tested to train the GP model with these less related individuals for the common traits, and the results were encouraging in 1 of the 2 populations.

Comparison of the populations

The 2 populations studied were similar in the sense that they were composed of biparental crosses with a resistant and emblematic grapevine variety as parents (Fig. 1). In both designs, the number

of individuals per cross was highly unbalanced, especially in the VS (Supplementary Fig. 1). We observed that the number of remaining SNP markers was higher in EDGARR than in Martell population (27,271 and 10,602, respectively), despite a higher number of reads per genotype for Martell (4.6 M) compared with EDGARR (4 M). This might be explained by the broader genetic diversity and less genetic relatedness of the resistant parents in the Martell population.

The size of the entire population for Martell was about twice the size of EDGARR. Nevertheless, PA observed was similar for both populations, with a comparable range and average. There were only 2 common traits between these populations: berry weight (SBER_W_g) and number of clusters (NB_CLUST_PLANT). Other traits were close, such as malic and tartaric acids, or total acidity, but they were not measured on the same entity (berry for EDGARR and must for Martell). Then, they were considered as different traits. PA for the number of clusters was extremely different in the 2 populations, with low PA for EDGARR (0.09) and high PA for Martell (0.78) (Fig. 3). This might be explained by the fact that for EDGARR, the number and length of shoots were not controlled. Then, the number of clusters is relative to the number of shoots and the fertility. This result is consistent with the difference in heritability values between these populations (Supplementary Table 3). For EDGARR, traits for sugar concentrations displayed low heritability and PA, probably because the sampling date was determined by a sugar threshold; thus, the genetic variability for these traits was minimized. These results illustrate the effect of vineyard management and measurement methodology on heritability values.

Factors affecting the PA

A major factor impacting PA was the presence of the cross effect in the final BLUP model (Supplementary Fig. 3). We found that traits with the cross effect had more differentiated genetic values per cross (defined as the sum of the cultivar and cross effects). Then, PA was automatically increased because we predicted both the average of a cross and the Mendelian sampling part (within a cross) (Würschum et al. 2017; Werner et al. 2020). This effect was highlighted by Werner et al. (2020), who measured PA per cross and for several crosses. However, we could not use a single cross as training or validation population, because we did not have enough genotypes and cross sizes were unbalanced. Brault, Segura, et al. (2022) compiled predicted genotypic values per cross and calculated the PA of GP for each cross and PA for predicting the cross means. But again, we had too few individuals to accurately measure PA for each cross.

As expected, the heritability values were overall correlated with PA values for both populations.

Across-population GP was competitive with within-population GP for the EDGARR population (Fig. 5). This was unexpected since the TS used in across-population scenario was phenotyped in the field and during different years compared with EDGARR population, phenotyped in pots. In the Martell population, PA values were higher in the within-population scenario, and differences between within- and across-population GP were higher compared with EDGARR population. However, we observed that PA values in across-population were higher for EDGARR than for Martell for SBER_W_g (Fig. 5), while TS sizes were constant. Then, this difference in PA in the across-population scenario could be due to the differences in genetic relatedness between TS and VS or by the phenotyping environment. Indeed, the genetic relatedness between TS and VS was higher for EDGARR than for Martell (Supplementary Fig. 7). The diversity panel and the half-diallel

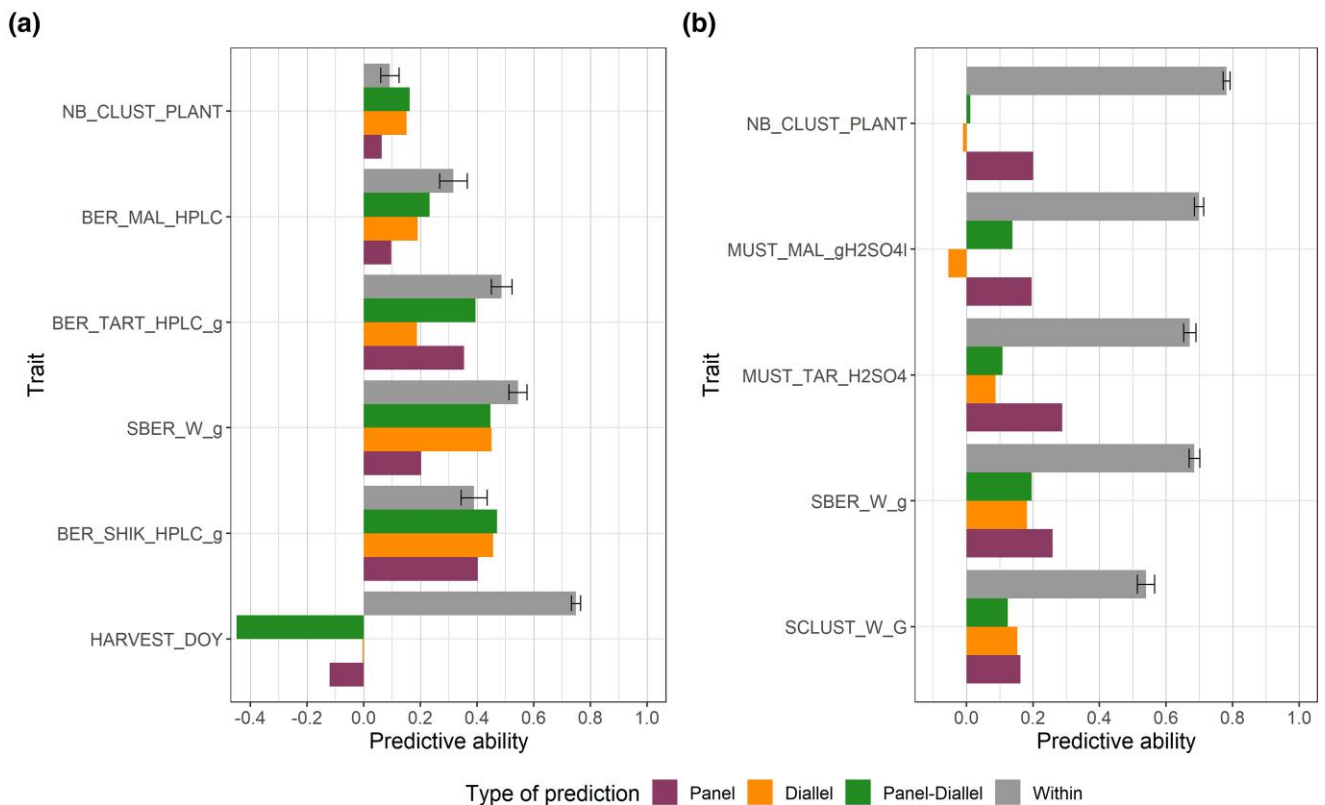


Fig. 5. Comparison of the PA for various TSs. a) EDGARR population and b) Martell population.

were planted about 20 km apart from the EDGARR population and about 400 km apart from the Martell population. Our results suggest that the geographic proximity of TS and VS and genetic relatedness could have more impact on PA than TS size. Since the genetic relatedness between Martell or EDGARR varieties and half-diallel or diversity panel vary greatly (Supplementary Fig. 7), we would expect further improvement of prediction accuracy with TS optimization.

In the Martell population, we studied semiquantitative traits, which displayed slightly smaller PA than other traits. We considered such traits as normal traits, even if the assumption of normality was strongly violated. Recently, Azevedo *et al.* (2023) showed that using a linear mixed model for GP of ordinal traits was robust but suboptimal. They advised using Bayesian Ordinal Regression Models, even though it is computationally demanding.

Future breeding programs

These breeding programs aimed to save time and maximize the genetic relatedness between training and VSs. First, individuals were filtered by MAS for disease resistance and hermaphroditism (Fig. 1). The discarded individuals were quickly planted in pots to be phenotyped and serve as the TS, while genotypic values could be predicted for the VS, using GP. Such a breeding program relies on 2 strong hypotheses: (1) phenotypes do not display a high genotype-by-environment ($G \times E$) interaction between pots and the field, and (2) genetic relatedness is a major parameter of PA. Indeed, if we observe a strong $G \times E$ interaction, the ranking of individuals between pot and field will likely vary, hampering an accurate selection of the best individuals. To some extent, this was tested in the across-population scenario and PA values were nearly as high as they were in the within-population scenario for some

traits for EDGARR population. This hypothesis should be further investigated for more traits and scenarios. For the second hypothesis, if genetic relatedness was already known to affect PA, its magnitude remains unknown, especially in this study where the VS was composed of interspecific varieties, while phenotypic data were only available for *V. vinifera* varieties. This is the first time that GP has been applied with such different genetic backgrounds between the training and the VSs. We tested using completely different populations to train the model, and results were encouraging for most traits for EDGARR population, while PA values were smaller in across-population for Martell population.

For across-population GP, we showed that LASSO was more often better than GBLUP, compared with the within-population scenario (Supplementary Fig. 6). This observation was also done in another study on grapevine (Brault, Segura, *et al.* 2022).

In this study, we used prediction models harnessing only the additive variance, while we expect for a heterozygous species such as grapevine that dominance could have a sizable proportion of genetic variance for some traits. In the future, having models taking into account all variance components would be more appropriate in the breeding process, as it was done by Wolfe *et al.* (2021).

Phenotyping environment

In our design of experiment, there was no repetition of a given genotype for a given year. Despite this, we could have medium to high heritability values depending on the trait. These values must be taken with caution, as variance components are likely not well estimated with this design.

Potted own-rooted grapevine phenotypes are likely to differ compared with field phenotypes. However, we have not found studies that compared both different varieties and traits related to the harvest. Most studies on pots or greenhouses were focused

on disease resistance or drought tolerance. If this kind of breeding is chosen for the future, one should measure the $G \times E$ interactions beforehand.

Grapevine ideotype

For EDGARR, a variety for Rosé wine was sought, with a little color, while for Martell, a variety for Cognac production was sought, with a white berry color and high yield. Beyond those criteria, both projects were aiming to counterbalance the effects of climate change on berry composition, namely higher alcohol degree, lower acidity, and shorter growth period (Rienth et al. 2016, 2021; de Cortázar-Atauri et al. 2017; Parker et al. 2020; Bécart et al. 2022). These traits interact with each other's. Selecting varieties that are ripening later (i.e. at the beginning of autumn in the Northern hemisphere) will experience lower temperatures during ripening, which would slow the degradation of malic acid and the accumulation of sugar (van Leeuwen et al. 2019). Ideotypes are now integrating traits related to the wine product, climate change, disease resistance, and more generally to production (yield, ability to produce wine). Other traits not directly in the ideotype would also be important, such as the resistance to black-rot *Guignardia bidwellii*, to *millerandage* and to *coulure* (poor fruit set). Besides, one may want to select individuals with medium performance across the traits or to correct the default of current grape varieties. The last solution is possible only if musts are blended.

As many other traits could not be included in the ideotype because of the difficulty of phenotyping, one must ensure that the selection intensity is not too high. Thus, enough individuals with genetic diversity must be kept to be phenotyped for costly traits such as wine aromas later in the breeding program.

Another solution for grapevine breeding would be to predict the best crosses to realize, based on the cross mean and variance prediction. The proof-of-concept for cross mean was already done in grapevine (Brault, Segura, et al. 2022), but it was not applied in a breeding context. Predicting cross variance would allow to select crosses that would result in extreme offspring phenotypes (Neyhart and Smith 2019; Wolfe et al. 2021).

In contrast to other crops, the grapevine ideotype is likely to include traits for which an optimum value is sought. That is why we used deregressed genetic values so that the range of values for these traits remains meaningful to breeders. However, such double deregression as we did here could hamper the prediction quality. For the mixed model, we could have used BLUEs instead of BLUPs, but the design of experiment was too unbalanced, especially for the number of individuals per cross.

Conclusion

This study provided the first insights on how GP could be integrated into grapevine breeding programs. The comparison of 2 breeding programs helped us identify factors affecting the prediction accuracy and determining the best conditions for applying GP, notably the training population environment and phenotypic reliability. For the first time in grapevine, a multitrait selection index was used based on predicted genotypic values to help select the best cultivars.

Data availability

Data and code to reproduce the results are available at doi:10.57745/G8PXJE (Segura et al. 2023). Genomic and phenotypic data for half-diallel and diversity panel populations are available at doi:10.15454/PNQUQU (Brault et al. 2023).

Supplemental material available at G3 online.

Acknowledgments

Keygene N.V. owns patents and patent applications protecting its Sequence-Based Genotyping technologies, patents WO2006/137733, WO2007/073165, and WO2007/114693.

Funding

Both Martell and EDGARR programs used for this publication were made possible through funding provided by CASDAR, the Special Account for Agricultural and Rural Development, from Ministère de l'Agriculture et de la souveraineté alimentaire.

Conflicts of interest

The author(s) declare no conflicts of interest.

Author contributions

C.B. performed data curation and statistical analysis and wrote the manuscript with inputs from all authors; V.S. and T.F. contributed to the design; V.S., L.L.C., and T.F. aided in interpreting the results and worked on the manuscript; L.L.C. designed and directed the project; M.R. performed genomic experiments and bioinformatics analysis; P.L., V.B., M.B., M.F., L.G., L.C., M.-A.D., C.R., and L.L.C. carried out phenotyping and data management; L.L.C., N.P., C.C., and G.M. were in charge of the EDGARR program; L.L.C., M.F., and S.J. were in charge of the Martell program.

Literature cited

- Alonso-Blanco C, Méndez-Vigo B. 2014. Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. *Curr Opin Plant Biol.* 18:37–43. doi:10.1016/j.pbi.2014.01.002.
- Andrade LRB, de Sousa MB, Oliveira EJ, Resende MDV, de Azevedo CF. 2019. Cassava yield traits predicted by genomic selection methods. *PLoS One.* 14(11):e0224920. doi:10.1371/journal.pone.0224920.
- Andrews S. 2016. FastQC a quality control tool for high throughput sequence data [WWW Document]. <https://www.bioinformatics.babraham.ac.uk/projects/fastqc/> [accessed 2023 Jun 12].
- Azevedo CF, Ferrão LFV, Benevenuto J, Resende MDV, de Nascimento Moyses, Nascimento ACC, Munoz PR. 2023. Using visual scores for genomic prediction of complex traits in breeding programs. *Theoretical and Applied Genetics.* 137(1):1–9. doi:10.1007/s00122-023-04512-w.
- Bates D, Mächler M, Bolker B, Walker S. 2014. Fitting linear mixed-effects models using lme4. *arXiv:1406.5823*, preprint: not peer reviewed.
- Beavis WD, Smith OS, Grant D, Fincher R. 1994. Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. *Crop Sci.* 34(4):882–896. doi:10.2135/cropsci1994.0011183X003400040010x.
- Bécart V, Lacroix R, Puech C, de Cortázar-Atauri IG. 2022. Assessment of changes in Grenache grapevine maturity in a Mediterranean context over the last half-century. *OENO One.* 56(1):53–72. doi:10.20870/oeno-one.2022.56.1.4727.
- Bernardo R. 1994. Prediction of maize single-cross performance using RFLPs and information from related hybrids. *Crop Sci.* 34(1): 20–25. doi:10.2135/cropsci1994.0011183X003400010003x.
- Brault C, Doligez A, Cunff L, Coupel-Ledru A, Simonneau T, Chiquet J, This P, Flutre T. 2021. Harnessing multivariate, penalized

- regression methods for genomic prediction and QTL detection of drought-related traits in grapevine. *G3 (Bethesda) Genes|Genomes|Genetics*. 11(9):jkab248. doi:[10.1093/g3journal/jkab248](https://doi.org/10.1093/g3journal/jkab248).
- Brault C, Lazerges J, Doligez A, Thomas M, Ecartot M, Roumet P, Bertrand Y, Berger G, Pons T, François P, et al. 2022. Interest of phenomic prediction as an alternative to genomic prediction in grapevine. *Plant Methods*. 18(1):108. doi:[10.1186/s13007-022-00940-9](https://doi.org/10.1186/s13007-022-00940-9).
- Brault C, Segura V, This P, Le Cunff L, Flutre T, François P, Pons T, Péros J-P, Doligez A. 2022. Across-population genomic prediction in grapevine opens up promising prospects for breeding. *Hortic Res*. 9:uhac041. doi:[10.1093/hr/uhac041](https://doi.org/10.1093/hr/uhac041).
- Brault C, Segura V, This P, Le Cunff L, Flutre T, François P, Pons T, Péros J-P, Doligez A. 2023. Genomic prediction between half-diallel and diversity panel populations in grapevine - phenotypic data and R scripts. *Recherche Data Gouv*. doi:[10.15454/PNQQUQ](https://doi.org/10.15454/PNQQUQ).
- Browning BL, Zhou Y, Browning SR. 2018. A one-penny imputed genome from next-generation reference panels. *Am J Hum Genet*. 103(3):338–348. doi:[10.1016/j.ajhg.2018.07.015](https://doi.org/10.1016/j.ajhg.2018.07.015).
- Crossa J, Pérez-Rodríguez P, Cuevas J, Montesinos-López O, Jarquín D, de Los Campos G, Burgueño J, González-Camacho JM, Pérez-Elizalde S, Beyene Y, et al. 2017. Genomic selection in plant breeding: methods, models, and perspectives. *Trends Plant Sci*. 22(11):961–975. doi:[10.1016/j.tplants.2017.08.011](https://doi.org/10.1016/j.tplants.2017.08.011).
- de Carvalho Rocha JRAS, Machado JC, Carneiro PCS. 2018. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. *GCB Bioenergy*. 10(1):52–60. doi:[10.1111/gcbb.12443](https://doi.org/10.1111/gcbb.12443).
- de Cortázar-Atauri IG, Duchêne E, Destrac-Irvine A, Barbeau G, de Rességuier L, Lacombe T, Parker AK, Saurin N, van Leeuwen C. 2017. Grapevine phenology in France: from past observations to future evolutions in the context of climate change. *OENO One*. 51(2):115–126. doi:[10.20870/oenone.2017.51.2.1622](https://doi.org/10.20870/oenone.2017.51.2.1622).
- DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, et al. 2011. A framework for variation discovery and genotyping using next-generation DNA sequencing data. *Nat Genet*. 43(5):491–498. doi:[10.1038/ng.806](https://doi.org/10.1038/ng.806).
- Eibach R, Zyprian E, Welter L, Töpfer R. 2007. The use of molecular markers for pyramiding resistance genes in grapevine breeding. *J Grapevine Res* 46(3):120–124. doi:[10.17660/ActaHortic.2009.827.96](https://doi.org/10.17660/ActaHortic.2009.827.96)
- Elshire RJ, Glaubitz JC, Sun Q, Poland JA, Kawamoto K, Buckler ES, Mitchell SE. 2011. A robust, simple genotyping-by-sequencing (GBS) approach for high diversity species. *PLoS One*. 6(5):e19379. doi:[10.1371/journal.pone.0019379](https://doi.org/10.1371/journal.pone.0019379).
- Endelman JB. 2011. Ridge regression and other kernels for genomic selection with R package rrBLUP. *Plant Genome*. 4(3):250–255. doi:[10.3835/plantgenome2011.08.0024](https://doi.org/10.3835/plantgenome2011.08.0024).
- Flutre T, Le Cunff L, Fodor A, Launay A, Romieu C, Berger G, Bertrand Y, Terrier N, Beccavin I, Bouckennooghe V, et al. 2022. A genome-wide association and prediction study in grapevine deciphers the genetic architecture of multiple traits and identifies genes under many new QTLs. *G3 (Bethesda) Genes|Genomes|Genetics*. 12(7):jkac103. doi:[10.1093/g3journal/jkac103](https://doi.org/10.1093/g3journal/jkac103).
- Fodor A, Segura V, Denis M, Neuenschwander S, Fournier-Level A, Chatelet P, Homa FAA, Lacombe T, This P, Le Cunff L. 2014. Genome-wide prediction methods in highly diverse and heterozygous species: proof-of-concept through simulation in grapevine. *PLoS One*. 9(11):e110436. doi:[10.1371/journal.pone.0110436](https://doi.org/10.1371/journal.pone.0110436).
- Friedman J, Hastie T, Tibshirani R. 2010. Regularization paths for generalized linear models via coordinate descent. *J Stat Softw*. 33(1):1–22. doi:[10.18637/jss.v033.i01](https://doi.org/10.18637/jss.v033.i01).
- R2D2 Consortium, Fugerey-Scarbel A, Bastien C, Dupont-Nivet M, Lemarié S. 2021. Why and how to switch to genomic selection: lessons from plant and animal breeding experience. *Front Genet*. 12:629737. doi:[10.3389/fgene.2021.629737](https://doi.org/10.3389/fgene.2021.629737).
- Garrick DJ, Taylor JF, Fernando RL. 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. *Genet Sel Evol*. 41(1):55. doi:[10.1186/1297-9686-41-55](https://doi.org/10.1186/1297-9686-41-55).
- Heffner EL, Sorrells ME, Jannink J-L. 2009. Genomic selection for crop improvement. *Crop Sci*. 49(1):1–12. doi:[10.2135/cropsci2008.08.0512](https://doi.org/10.2135/cropsci2008.08.0512).
- Li H. 2013. Aligning sequence reads, clone sequences and assembly contigs with BWA-MEM. *arXiv13033997 Q-Bio*, preprint: not peer reviewed.
- Lush JL. 1937. *Animal Breeding Plans*. Iowa State College Press: Collegiate Press. <https://www.cabidigitallibrary.org/doi/full/10.5555/19440100562>.
- Mahalanobis PC. 1936. On the generalized distance in statistics. *Indian J Stat*. 2:49–55. <https://www.jstor.org/stable/4872335>.
- Martin M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet J*. 17(1):10–12. doi:[10.14806/ej.17.1.200](https://doi.org/10.14806/ej.17.1.200).
- Meuwissen T, Hayes B, Goddard M. 2001. Prediction of total genetic value using genome-wide dense marker maps. *Genetics*. 157(4):1819–1829. doi:[10.1093/genetics/157.4.1819](https://doi.org/10.1093/genetics/157.4.1819).
- Meuwissen T, Hayes B, Goddard M. 2016. Genomic selection: a paradigm shift in animal breeding. *Anim Front*. 6(1):6–14. doi:[10.2527/af.2016-0002](https://doi.org/10.2527/af.2016-0002).
- Migicovsky Z, Sawler J, Gardner KM, Aradhya MK, Prins BH, Schwaninger HR, Bustamante CD, Buckler ES, Zhong G-Y, Brown PJ, et al. 2017. Patterns of genomic and phenomic diversity in wine and table grapes. *Hortic Res*. 4(1):17035. doi:[10.1038/hortres.2017.35](https://doi.org/10.1038/hortres.2017.35).
- Neyhart JL, Smith KP. 2019. Validating genomewide predictions of genetic variance in a contemporary breeding program. *Crop Sci*. 59(3):1062–1072. doi:[10.2135/cropsci2018.11.0716](https://doi.org/10.2135/cropsci2018.11.0716).
- Nicolas SD, Péros J-P, Lacombe T, Launay A, Le Paslier M-C, Bérard A, Mangin B, Valière S, Martins F, Le Cunff L, 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(1):74. doi:[10.1186/s12870-016-0754-z](https://doi.org/10.1186/s12870-016-0754-z).
- Olivoto T, Nardino M. 2021. MGIDI: toward an effective multivariate selection in biological experiments. *Bioinformatics*. 37(10):1383–1389. doi:[10.1093/bioinformatics/btaa981](https://doi.org/10.1093/bioinformatics/btaa981).
- Parker AK, de Cortázar-Atauri IG, Gény L, Spring J-L, Destrac A, Schultz H, Molitor D, Lacombe T, Graça A, Monamy C, et al. 2020. Temperature-based grapevine sugar ripeness modelling for a wide range of *Vitis vinifera* L. cultivars. *Agric For Meteorol*. 285–286:107902. doi:[10.1016/j.agrformet.2020.107902](https://doi.org/10.1016/j.agrformet.2020.107902).
- Reynolds A. 2015. *Grapevine Breeding Programs for the Wine Industry*. Elsevier Science. p. 1–466. <https://www.sciencedirect.com/book/9781782420750/grapevine-breeding-programs-for-the-wine-industry>.
- Rienth M, Torregrosa L, Sarah G, Ardisson M, Brillouet J-M, Romieu C. 2016. Temperature desynchronizes sugar and organic acid metabolism in ripening grapevine fruits and remodels their transcriptome. *BMC Plant Biol*. 16(1):164. doi:[10.1186/s12870-016-0850-0](https://doi.org/10.1186/s12870-016-0850-0).
- Rienth M, Vigneron N, Darriet P, Sweetman C, Burbidge C, Bonghi C, Walker RP, Famiani F, Castellarin SD. 2021. Grape berry secondary metabolites and their modulation by abiotic factors in a climate change scenario—a review. *Front Plant Sci*. 12:643258. doi:[10.3389/fpls.2021.643258](https://doi.org/10.3389/fpls.2021.643258).
- Santos JA, Fraga H, Malheiro AC, Moutinho-Pereira J, Dinis L-T, Correia C, Moriondo M, Leolini L, Dibari C, Costafreda-Aumedes S, et al. 2020. A review of the potential climate change impacts and adaptation options for European viticulture. *Appl Sci*. 10(9):3092. doi:[10.3390/app10093092](https://doi.org/10.3390/app10093092).

- Schneider C, Onimus C, Prado E, Dumas V, Wiedemann-Merdinoglu S, Dorne MA, Lacombe MC, Piron MC, Umar-Faruk A, Duchêne E, et al. 2019. INRA-ResDur: the French grapevine breeding programme for durable resistance to downy and powdery mildew. *Acta Hort.* 1248:207–214. doi:[10.17660/ActaHortic.2019.1248.30](https://doi.org/10.17660/ActaHortic.2019.1248.30).
- Segura V, Brault C, Le Cunff L. 2023. Grapevine breeding programs and selection index - script and data. *Recherche Data Gov.* doi:[10.57745/G8PXEJ](https://doi.org/10.57745/G8PXEJ).
- Smith HF. 1936. A discriminant function for plant selection. *Ann Eugen.* 7(3):240–250. doi:[10.1111/j.1469-1809.1936.tb02143.x](https://doi.org/10.1111/j.1469-1809.1936.tb02143.x).
- Tello J, Roux C, Chouiki H, Laucou V, Sarah G, Weber A, Santoni S, Flutre T, Pons T, This P, et al. 2019. A novel high-density grapevine (*Vitis vinifera* L.) integrated linkage map using GBS in a half-diallel population. *Theor Appl Genet.* 132(8):2237–2252. doi:[10.1007/s00122-019-03351-y](https://doi.org/10.1007/s00122-019-03351-y).
- Tibshirani R. 1996. Regression shrinkage and selection via the Lasso. *J R Stat Soc Ser B Methodol.* 58:267–288. <https://www.jstor.org/stable/2346178>.
- Töpfer R, Hausmann L, Harst M, Maul E, Zyprian E, Eibach R. 2010. New horizons for grapevine breeding. In Flachowsky H, Hanke MV (eds) *Fruit, Vegetable and Cereal Science and Biotechnology*, vol 5. *Methods in Temperate Fruit Breeding*. Isleworth: Global Science Books. p. 79–100.
- Töpfer R, Trapp O. 2022. A cool climate perspective on grapevine breeding: climate change and sustainability are driving forces for changing varieties in a traditional market. *Theor Appl Genet.* 135(11):3947–3960. doi:[10.1007/s00122-022-04077-0](https://doi.org/10.1007/s00122-022-04077-0).
- van Leeuwen C, Destrac-Irvine A, Dubernet M, Duchêne E, Gowdy M, Marguerit E, Pieri P, Parker AK, de Ressaiguier L, Ollat N. 2019. An update on the impact of climate change in viticulture and potential adaptations. *Agronomy.* 9(9):514. doi:[10.3390/agronomy9090514](https://doi.org/10.3390/agronomy9090514).
- Velt A, Frommer B, Blanc S, Holtgräwe D, Duchêne É, Dumas V, Grimplet J, Hugueney P, Kim C, Lahaye M, et al. 2023. An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. G3 (Bethesda): Genes, Genomes, Genetics. 13:jkad067. doi:[10.1093/g3journal/jkad067](https://doi.org/10.1093/g3journal/jkad067).
- Viana AP, Resende MDV, de Riaz S, Walker MA. 2016. Genome selection in fruit breeding: application to table grapes. *Scientia Agricola.* 73(2):142–149. doi:[10.1590/0103-9016-2014-0323](https://doi.org/10.1590/0103-9016-2014-0323).
- Werner CR, Gaynor RC, Gorjanc G, Hickey JM, Kox T, Abbadi A, Leckband G, Snowdon RJ, Stahl A. 2020. How population structure impacts genomic selection accuracy in cross-validation: implications for practical breeding. *Front Plant Sci.* 11:592977. doi:[10.3389/fpls.2020.592977](https://doi.org/10.3389/fpls.2020.592977).
- Wolfe MD, Chan AW, Kulakow P, Rabbi I, Jannink J-L. 2021. Genomic mating in outbred species: predicting cross usefulness with additive and total genetic covariance matrices. *Genetics.* 219(3): iyab122. doi:[10.1093/genetics/iyab122](https://doi.org/10.1093/genetics/iyab122).
- Würschum T, Maurer HP, Weissmann S, Hahn V, Leiser WL. 2017. Accuracy of within- and among-family genomic prediction in triticale. *Plant Breed.* 136(2):230–236. doi:[10.1111/pbr.12465](https://doi.org/10.1111/pbr.12465).
- Xu S. 2003. Theoretical basis of the Beavis effect. *Genetics.* 165(4): 2259–2268. doi:[10.1093/genetics/165.4.2259](https://doi.org/10.1093/genetics/165.4.2259).

Editor: R. Wisser