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 Bouckenooghe[1,2], Nathalie Pouzalgues[4], Constance Cunty[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]*

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

[2] Institut Français de la vigne et du vin, 30240, Le Grau du Roi, France.

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

[4] Centre du Rosé, 70, Av. du Président Wilson, 83550 Vidauban, France

[5] Conservatoire du Vignoble Charentais, Trepsec, 16370 Cherves-Richemont, France

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

Corresponding authors*: loic.lecunff@vignevin.com, charlotte.brault@live.com

Abstract

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 to 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 two existing grapevine breeding programs: *Rosé* wine and *Cognac* production. In these programs, several families were created through crosses of emblematic and inter-specific resistant varieties to powdery and downy mildews. 30 traits were evaluated for each program, using two genomic prediction methods: GBLUP (Genomic Best Linear Unbiased Predictor) and LASSO (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.

Introduction

Plant breeding has been a key lever to adapt varieties to human use and the environment. The genetic gain obtained after one 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 to 25 years, when accounting for phenotyping new varieties (Töpfer and Trapp, 2022). Thus, grapevine breeding is critically long and hereafter the genetic gain is reduced. Because of its perennial nature, grapevine needs to be adapted to challenging conditions, in a 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 two major resistance genes for each disease. The whole selection process lasted around 15 to 20 years (Reynolds and TBX, 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 to early detect seedlings with all resistance genes. However, most quantitative traits involved in adaptation to climate change are under a complex genetic determinism, with possibly thousands of genes involved (Alonso-Blanco and Méndez-Vigo, 2014; Flutre et al., 2022). In that case, QTL detection results in many small effects often overestimated and that are not transferable through MAS to breeding (Beavis et al., 1994; Crossa et al., 2017; Meuwissen et al., 2016; Xu, 2003).

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 population (TS). Then, these effects are applied in a validation set population (VS), on which only genotypes are available (Heffner et al., 2009). GS has been widely applied to animal and plant breeding, with some scarce examples of applications in grapevine (Brault et al., 2021, 2022b; Flutre et al., 2022; Fodor et al., 2014; Migicovsky et al., 2017; Viana et al., 2016). Especially, GS has only been applied in a

research context, with varieties not intended for breeding. GS allows to save time in the breeding programs but it offers other interests (Consortium et al., 2021). Indeed, using GS allows testing of more crosses and offspring because no phenotyping is needed. Then, the selection intensity is increased, 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 (Brault et al., 2022b). 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 multi-trait 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 (Olivoto and Nardino, 2021; Rocha et al., 2018). 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 and TBX, 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 and TBX, 2015).

This article describes and proposes an application of GS to two breeding programs of grapevine varieties. These two 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, then we applied genomic prediction within the training set to estimate predictive ability. Finally, we used multi-trait selection index to select the most promising individuals from predicted genotypic values.

Material and methods

Design of experiment

Two breeding programs 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

95 in France; and the EDGARR breeding programs, conducted by the *Center for Rosé,* INRAE, and IFV in
96 France for producing Rosé wine. Both programs included crosses between varieties emblematic of the
97 region and varieties with polygenic resistance to powdery and downy mildews (inter-specific hybrids).
98 In both programs, after MAS, unselected individuals from the crosses were planted in a pot to
99 constitute the TS, for genotyping and phenotyping while selected individuals were only genotyped and
100 constitute the VS, except for a few families only present in the VS (Figure S1).

101 The Martell program included four famous grape varieties (Monbadon, Montils, Rayon d'Or, and Vidal
102 36), and the EDGARR experiment included two famous grape varieties (Cinsaut and Vermentino). The

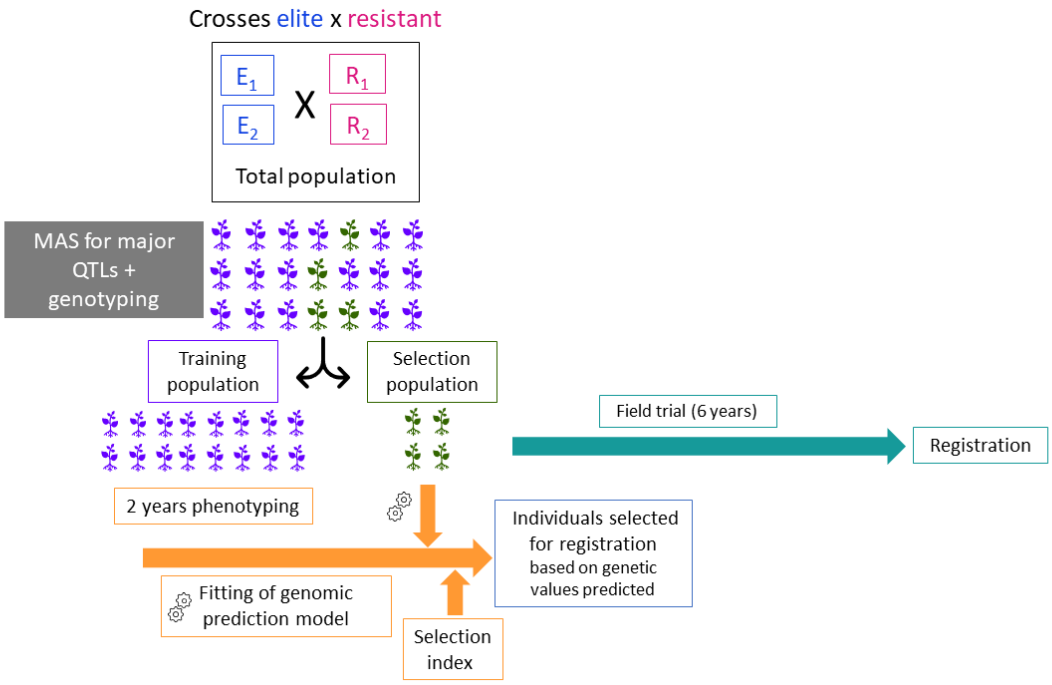

Figure 1 Design of experiment for EDGARR and Martell breeding programs

103

104 genetic relatedness between the individuals of the TS and the VS could be full-sibs, half-sibs, or no
105 genetic relationship. A major difference between these programs was the number of genotypes. In the
106 Martell program, there were 347 and 277 individuals in the TS and VS, respectively. In the EDGARR
107 program, there were 193 and 132 individuals in the TS and VS, respectively.

108

109 Genomic data analysis

110 The same genotyping approach was used in both programs. Genotyping was done using the
111 genotyping-by-sequencing technology, using the *ApeKI* restriction enzyme (Elshire et al., 2011).
112 Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping
113 technologies.

For EDGARR and Martell programs, SNP markers with more than 10% missing data and with less than 20 reads were discarded, producing 27,271 and 10,602 remaining SNPs, respectively. These markers were mapped to the updated version of the reference genome, PN40024.v4.2 (Velt et al., 2023). 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 one repetition per genotype, except for a small number of repeated controls (Cinsaut, Vermentino, Grenache and Syrah, with 5 or 6 repetitions).

In this population, 30 traits were phenotyped for two years (2018 and 2019), and 5 additional traits were phenotyped for one year. Traits were divided into five categories, namely acids with cis- and trans-coutaric acids, caftaric, ascorbic, hydroxycinnamic, malic, shikimic, tartaric acids, pH, and total acidity; color traits with blue, yellow and red absorbance, lightness, yellow and red indices, color intensity, tint and polymeric pigments at 420 and 520 nm; sugar traits with glucose and fructose; polyphenol traits with total polyphenol index, 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 can be found in table S2 and table S3. Clusters were sampled when the sugar content reached 22° brix (gram of saccharose / 100 g). Some traits were measured with two non-redundant 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 two years:

$y_{ijkl} = \mu + \underline{G_i} + \underline{C_j} + \underline{x} + \underline{y} + Y_k + \epsilon_{ijkl}$ (1) , with y_{ijkl} the phenotypic observation for a given genotype i, cross j, year k and repetition l (for controls), μ the intercept, $\underline{G_i}$ the random effect of the genotype j (nested in cross i), $\underline{C_i}$ the random effect of the cross, $\underline{x}$ and $\underline{y}$ the random effects for

coordinates of the plant in the trial, Y_k the fixed effect of the year (two levels), and ϵ_{ijk} the residuals, assumed normally distributed. 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). Variance components were estimated with restricted maximum likelihood on the selected model. The broad-sense heritability (H^2) was computed as:

$$H^2 = \frac{\sigma_G^2 + \sigma_C^2}{\sigma_G^2 + \sigma_C^2 + \frac{\sigma_x^2 + \sigma_y^2 + \sigma_\epsilon^2}{mean\ nb\ reps\ per\ trial}} \quad (2)$$

With $\sigma_G^2, \sigma_C^2, \sigma_x^2, \sigma_y^2, \sigma_\epsilon^2$ variances associated to genotype, cross, plot coordinates and residuals. Fitting information for all traits is available in table S4. If the genotype effect was not selected in the model, we re-fitted a simpler model with only the genotype effect as a random effect and we applied 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: $drgBLUP = \frac{BLUP}{1-\left(\frac{PEV}{\sigma_G^2}\right)}$ (3) (Andrade et al., 2019; Garrick et al., 2009), 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. For traits measured for one year, averaged phenotypic data per genotype was used.

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 training set 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 can be found in table S2 and table S3. 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 (1) and some other effects: $y_{ijk} = \mu + \underline{G_i} + \underline{C_j} + \underline{x} + \underline{y} + Y_k + \underline{Rpv3_i} + \underline{Ren3_i} + \underline{Run1_Rpv1_i} + \underline{M_{ijk}} + \epsilon_{ijk}$, with as supplementary effects, the resistance genes Rpv3, Ren3 and Run1_Rpv1, and M a factor indicating the presence of available vine spur (if one spur and one cane were present, the pruning was simple guyot). We used the same equation (2) for computing the heritability for Martell population.

176 ## Genomic prediction

177 The same pipeline of analysis was applied to both programs. First, genomic prediction (GP) was applied
178 to the training population for all traits available with K-fold cross-validation, repeated R=10 times, with
179 K=5. We implemented two genomic prediction methods: GBLUP with rrBLUP R package (Endelman,
180 2011) and the LASSO (Tibshirani, 1996), with glmnet R package (Friedman et al., 2010). GBLUP is more
181 adapted to a complex genetic architecture (many QTLs), while LASSO is more adapted to a simpler
182 genetic architecture. Predictive ability (PA) was estimated as Pearson's correlation between observed
183 and the predicted genotypic values. PA values were averaged across folds and cross-validation
184 repetitions and standard errors were calculated.

185 The best method among the two was chosen for each population and trait and used to predict the
186 genotypic values for the VS. The model was refit on the whole TS (without cross-validation) for
187 predicting genotypic values for the VS. These values were deregressed using equation (3). For the
188 LASSO, the deregressed values were obtained by fitting the Ordinary Least Square estimator for all
189 selected markers in the training set.

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

193 ## Selection index

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

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

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

209 ## Other phenotypic and genomic data

210 We used genomic and phenotypic data from two other grapevine populations. A half-diallel population
211 composed of 628 individuals from 10 bi-parental crosses where 5 parents were involved (Tello et al.,
212 2019), phenotyped between 2013 and 2017. The second population is a diversity panel population of
213 277 genotypes, chosen to represent the genetic diversity of *Vitis vinifera* (Nicolas et al., 2016), and
214 phenotyped between 2011 and 2012. Phenotypic and genomic data from these populations were
215 already analyzed for genomic prediction and QTL detection in previous studies (Brault et al., 2022b,
216 2022a; Flutre et al., 2022).

217 There were 6 and 5 common traits with EDGARR and Martell programs, respectively. For genomic data,
218 we performed a BLAST (Basic Local Alignment Research Tool) analysis on flanking sequences to find
219 out the marker positions corresponding to the last version (PN40024.v4) of the *Vitis vinifera* reference
220 genome (Velt et al., 2023). Then, we kept the common markers between each population and the
221 target one (number of left SNPs displayed in Table S5). We fitted a GP model using GBLUP and LASSO
222 for half-diallel, diversity panel, or both populations and kept the best method to predict genotypic
223 values of EDGARR and Martell populations. We measured the predictive ability and compared it to the
224 values from within-population GP.

225 # Results

226 ## Genetic structure

227 For the EDGARR program, 325 individuals have been genotyped for 19,228 SNP markers after filters.
228 For the Martell program, there were 624 individuals genotyped for 10,380 SNPs. A principal
229 component analysis (PCA) was conducted to explore the genetic structure of the population. We found
230 that families were well separated, located between their parents. Individuals in training and validation
231 sets displayed a clear overlap, except for some families only in the validation set (Figure S1). The PCA
232 analysis showed some outlier individuals, spotted with the Mahalanobis distance. For the EDGARR
233 population, we excluded 3 individuals, all from Cinsaut x 3421-F02-PL5 cross; for the Martell
234 population, we excluded 4 individuals from 4 crosses.

235 Overall, for both populations, the relative position of families seems to be driven by the inter-specific
236 resistant parents. This is likely because they show more genetic diversity compared to *V. vinifera*
237 varieties.

238

239

240

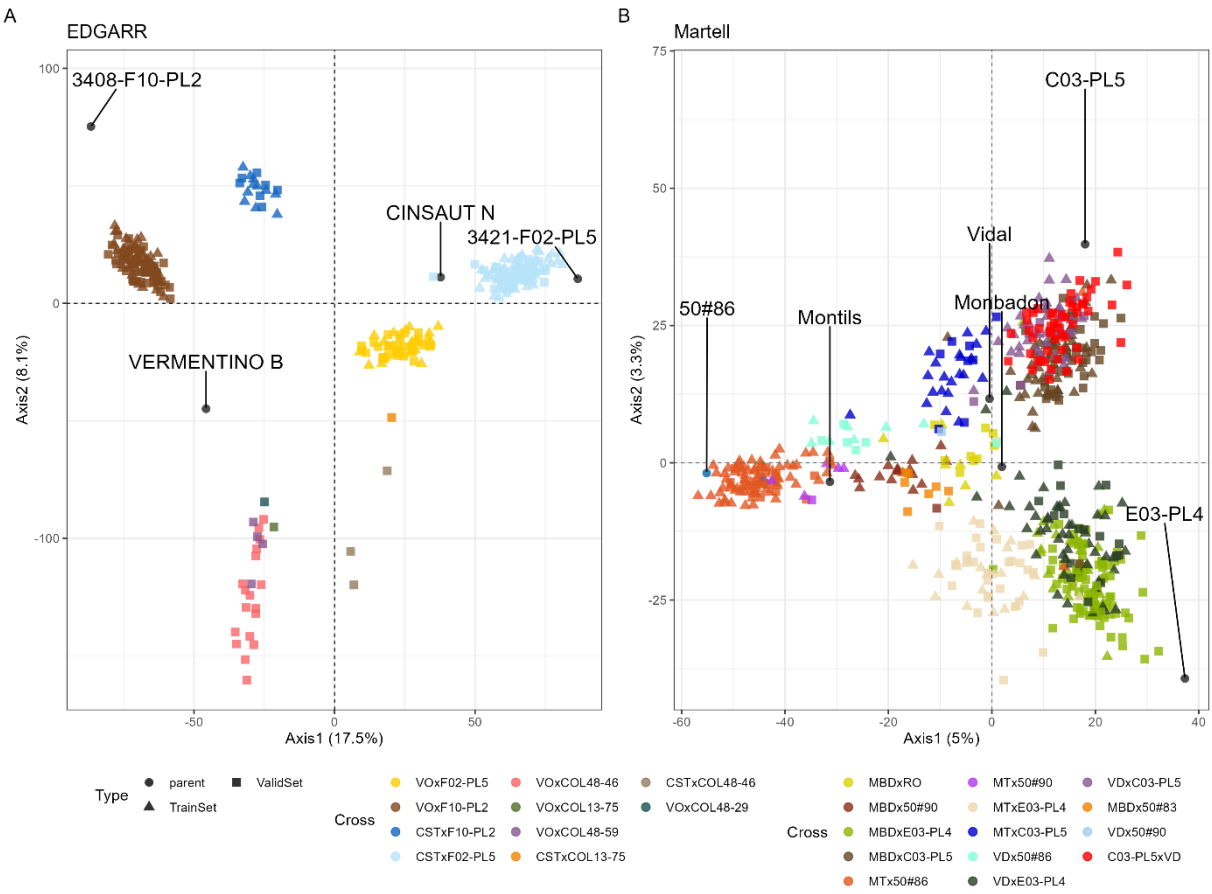

Figure 2 PCA of genetic markers for EDGARR (A) and Martell (B) populations.

Parents are labelled. The point shape corresponds to the type of individual, triangle: training set, square: validation set, points: parents of crosses. Cross names were abbreviated as follow: Vermentino (VO), Cinsaut (CST), Monbadon (MBD), Rayon d'Or (RO), Montils (MT) and Vidal (VD).

Phenotypic structure

For both populations, we included environmental cofactors, despite a small number of repetitions. We were able to estimate the effects of plot position, year, and resistance gene, depending on the trait (Table S4). Broad-sense heritability values displayed a wide range across all traits. 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 correlation between raw averaged phenotypic data and deregressed BLUPs was between 0.73 and 0.98 for the EDGARR population, and between 0.60 and 0.99 for the Martell population. The matrix of pairwise genotypic correlations between the traits showed for EDGARR, that traits related to color were correlated to each other. Overall, for the other traits, genotypic correlations were mostly low (data not shown).

255 The PCA analysis showed a mild phenotypic structure (Figure S6). For EDGARR, the structure was driven
256 by the inter-specific parents (F02-PL5 and F10-PL2) and by traits related to color, while for Martell, the
257 crosses were more separated from each other, and the differentiation was driven by acid and yield
258 traits.

259

260 ## Genomic prediction results

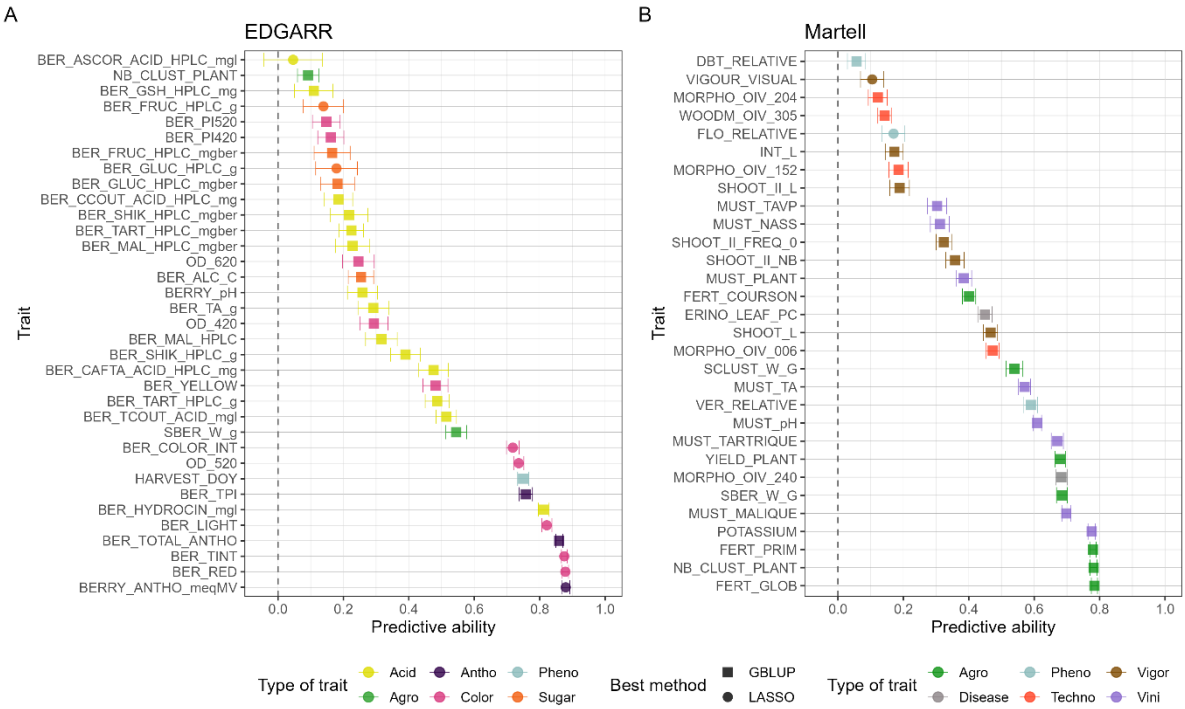

Figure 3 Predictive ability for all traits for EDGARR (A) and Martell (B) populations. Error bars correspond to standard errors calculated across cross-validation repetitions. For each trait, the best method among GBLUP and LASSO was selected.

261

262 Predictive abilities were comparable for both populations and covered a wide range of values between
263 0.04 to 0.87 (Figure 3). To avoid the effect of the genetic architecture on the predictive ability, we
264 chose the best method between GBLUP and LASSO. Overall, GBLUP provided a better PA than LASSO
265 for both populations, with an average of 0.41 and 0.34 for EDGARR and 0.44 and 0.39 for Martell, for
266 GBLUP and LASSO, respectively. For EDGARR, GBLUP yielded a higher PA than LASSO for 26 traits out
267 of 35, and 28 out of 30 for Martell. PA and heritability values were correlated for both populations,
268 with a correlation value of 0.60 for EDGARR, and 0.42 for Martell. The different trait categories were
269 quite evenly represented across the range of PA for both populations (Figure 3). However, traits for
270 which the cross effect was not kept in the mixed model (1), displayed a lower PA with an average
271 difference of 0.55 and 0.37 for EDGARR and Martell populations, respectively (Figure S7). We found

272 that the 4 traits measured on a semi-quantitative scale for Martell populations had a slightly lower PA
273 (difference of 0.26, a p-value of 0.044 using a Wilcoxon test). For EDGARR data, we could not fit a
274 mixed model for 6 traits, because they were phenotyped in a single year. For these traits, averaged
275 phenotypic data per genotype were used for GP. We found an average PA of 0.17 for these traits, up
276 to 0.51 for trans-coutaric acid and the GBLUP method.

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

282 We used a GLM with the LASSO method to predict categorical color for EDGARR population. The
283 accuracies ranged between 0.943 to 0.963, with an average of 0.952 in cross-validation.

284 Selection index

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

288 EDGARR selection index

289 For EDGARR population, the center of *Rosé* established a selection index to get varieties with more
290 acidity, less color, higher productivity, and adaptation to climate change. Finally, the corresponding
291 ideotype included 11 traits, 5 traits to be optimized (must tartaric, malic, total acidity, pH, alcohol
292 content), 2 traits to be minimized (color intensity and total polyphenol index) and 4 traits to be
293 maximized (berry tint, number of clusters, berry weight and harvest date) (Table S8). We used PA
294 values as weights associated with each trait. The MGIDI algorithm selected 3 factors, represented by 4
295 (tartaric acid, berry lightness, berry color, color intensity, harvest date), 3 (malic acid, total acidity, pH),
296 and 4 traits (alcohol content, total polyphenol index, number of cluster and berry weight), respectively
297 (Table S9). Vermentino was a parent of 12 out of 15 selected genotypes, and 8 individuals from the
298 same cross Vermentino x F10-PL2 were selected. Surprisingly, the resistant genotype F02-PL5 was not
299 selected as a parent of the first 15 genotypes. From the PCA analysis (Figure 4), it is clear that selected
300 individuals are phenotypically close to each other. The predicted berry color was white for 4 genotypes,
301 and the genotype with the lowest MGIDI was predicted white. Factors 1 and 2 contributed the most
302 to the MGIDI score for the selected genotypes, which means that they performed quite similarly for
303 factor 3. Some genotypes performed better for some factors, such as P869-F04 for factor 1, or P596-

304 A09 for factor 2, while others had a more balanced performance across factors, such as P249-F10
305 (Figure S12, Table S9).

306 ### Martell selection index

307 The ideotype for Martell included 11 traits, 7 traits to be maximized (global and primary fertility, yield,
308 tartaric acid, total must acidity, cluster weight and berry weight), 1 to be minimized (must pH), and 2
309 with an optimum value (must malic acid and ease of detachment of pedicel, OIV 240) (Table S8). We
310 excluded beforehand traits with a PA value lower than 0.5. The MGIDI algorithm selected 3 factors,
311 represented by 5 (total acidity, pH, yield, primary and global fertility), 4 (cluster weight, tartaric acid,
312 berry weight, malic acid) and 2 traits (ease of detachment of pedicel and potassium), respectively
313 (Table S9). Among the selected traits, some of them displayed high genetic correlations (positive or
314 negative). The average of the 15 genotypes selected followed the expected trend (increase or decrease
315 compared to the average of the population), for all the traits, except for single berry and cluster
316 weights (Table S9). Distributions of predicted genotypic values and the position of some parents and
317 selected genotypes are displayed in Figure S10. For 13 out of 15 genotypes selected, Monbadon and
318 C03-PL5 were one of the two parents (Table S11). As for EDGARR, factor 3 contributed less to the
319 MGIDI score and genotypes displayed various strengths or weaknesses for the factors. In particular,
320 the superior performance of E12-32G10 (ranked 1st) was due to factor 1 and 3, and E10-29D10 (ranked
321 7th) was only due to factors 2 and 3 (Figure S12, Table S9).

Figure 4. PCA of the genotypic values for the selection candidates for the traits in the selection index for the first two principal components. Variables are displayed in red, and genotypes are colored according to their cross. Selected individuals are labelled. A: EDGARR population, B: Martell population.

Across-population genomic prediction

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 (Figure 5). Overall, across-population PA values were much higher in EDGARR than in the Martell population. For EDGARR and

330 two traits (shikimic acid concentration and number of clusters), using data from the diversity panel
331 and the half-diallel led to a higher PA than using data from the same population.

332 For EDGARR, using both data from the diversity panel and the half-diallel led to higher PA, except for
333 the harvest date, for which a strong decrease was observed. For Martell, the diversity panel was the
334 best training population consistently for all traits.

335 We used for each trait and training set, the best method among GBLUP and LASSO. The results show
336 that GBLUP was the only method selected for within-population, while LASSO was the best method for
337 at least one trait for both populations.

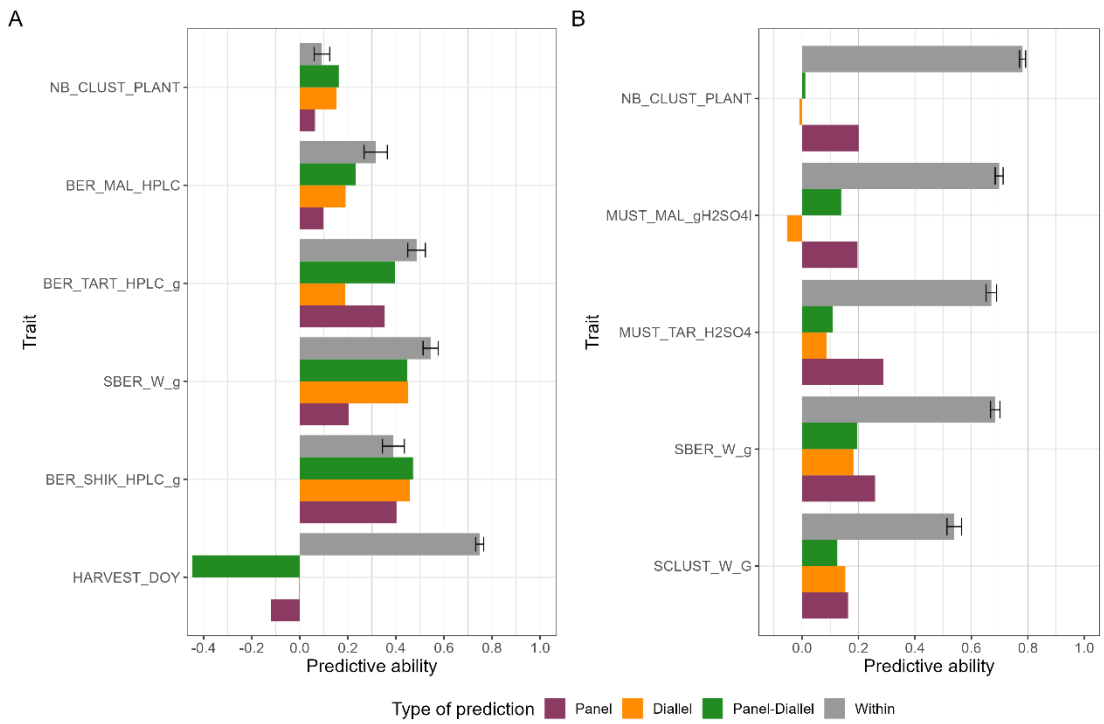

Figure 5 Comparison of the predictive ability for various training sets. A: EDGARR population, B: Martell population

338 ## Discussion

339 Our study comprised the analysis of 30 traits for two grapevine populations. Some of the individuals
340 were only genotyped, which allowed us to perform genomic predictions. We first tested the ability of
341 GP models to accurately predict the genotypic values in a within-population scenario. Then, we
342 proposed a selection index and selected the most relevant individuals according to it. The ideotype
343 was built in partnership with professional wine growers and was specific to each of the two wine
344 regions studied. To our knowledge, this is the first time a precise ideotype is described for grapevine.
345 Finally, some of the phenotyped traits were also available for other grapevine populations. We tested

346 to train the GP model with these less related individuals for the common traits and the results were
347 encouraging in one of the two populations.

348 Comparison of the populations

349 The two populations studied were similar in the sense that they were composed of bi-parental crosses
350 with a resistant and emblematic grapevine variety as parents (Figure 1). In both designs, the number
351 of individuals per cross was highly unbalanced, especially in the validation set (Figure S1). We observed
352 that number of remaining SNP markers was higher in EDGARR than in Martell population (27,271 and
353 10,602, respectively), despite a higher number of reads per genotype for Martell (4.6 M) compared to
354 EDGARR (4 M). This might be explained by the broader genetic diversity and less genetic relatedness
355 of the resistant parents in the Martell population.

356 The size of the entire population for Martell was about twice the size of EDGARR. Nevertheless, PA
357 observed were similar for both populations, with a comparable range and average. There were only
358 two common traits between these populations: berry weight (SBER_W_g) and number of clusters
359 (NB_CLUST_PLANT). Other traits were close, such as malic and tartaric acids, or total acidity, but they
360 were not measured on the same entity (berry for EDGARR and must for Martell). Then, they were
361 considered as different traits. PA for the number of clusters was extremely different in the two
362 populations, with low PA for EDGARR (0.09), and high PA for Martell (0.78) (Figure 3). This might be
363 explained by the fact that for EDGARR, the number and length of shoots were not controlled. Then,
364 the number of clusters is relative to the number of shoots and the fertility. This result is consistent
365 with the difference in heritability values between these populations (Figure S4). For EDGARR, traits for
366 sugar concentrations displayed low heritability and PA, probably because the sampling date was
367 determined by a sugar threshold, thus the genetic variability for these traits was minimized. These
368 results illustrate the effect of vineyard management and measurement methodology on heritability
369 values.

370 Factors affecting the predictive ability

371 A major factor impacting PA was the presence of the cross effect in the final BLUP model (Figure S7).
372 We found that traits with the cross effect had more differentiated genetic values per cross. Then, PA
373 was automatically increased because we predicted both the average of a cross and the Mendelian
374 sampling part (within a cross) (Werner et al., 2020; Würschum et al., 2017). This effect was highlighted
375 by Werner et al., (2020), who measured PA per cross and for several crosses. However, we could not
376 use a single cross as training or validation population, because we did not have enough genotypes and
377 cross sizes were unbalanced. Brault et al. (2022b), compiled predicted genotypic values per cross and

378 calculated the PA of GP for each cross and PA for predicting the cross means. But again, we had too
379 few individuals to accurately measure PA for each cross.

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

381 Across-population GP was competitive with within-population GP for the EDGARR population (Figure
382 5). This was unexpected since the TS used in across-population scenario was phenotyped in the field
383 and during different years compared to EDGARR population, phenotyped in pots. In the Martell
384 population, PA values were higher in the within-population scenario, and differences between within
385 and across-population GP was higher compared to EDGARR population. However, we observed that
386 PA values in across-population were higher for EDGARR than for Martell for SBER_W_g (Figure 5), while
387 TS sizes were constant. Then, this difference in PA in the across-population scenario could be due to
388 the differences in genetic relatedness between TS and VS, or by the phenotyping environment. The
389 diversity panel and the half-diallel were planted about 20 km apart from the EDGARR population, and
390 about 400 km apart from the Martell population. Our results suggest that the geographic proximity of
391 TS and VS could have more impact on PA than genetic relatedness or TS size.

392 In the Martell population, we studied semi-quantitative traits, which displayed slightly smaller PA than
393 other traits (Figure S7). We considered such traits as normal traits, even if the assumption of normality
394 was strongly violated. Recently, (Azevedo et al., 2023) showed that using a linear mixed model for GP
395 of ordinal traits was robust but sub-optimal. They advised using Bayesian Ordinal Regression Models,
396 even though it is computationally demanding.

397 Future breeding programs

398 These breeding programs aimed to save time and maximize the genetic relatedness between training
399 and validation sets. First, individuals were filtered by MAS for disease resistance and hermaphroditism
400 (Figure 1). The discarded individuals were quickly planted in pots to be phenotyped and serve as the
401 TS, while genotypic values could be predicted for the VS, using GP. Such a breeding program relies on
402 two strong hypotheses: i) phenotypes do not display a high genotype-by-environment (GxE)
403 interaction between pots and the field, and ii) genetic relatedness is a major parameter of PA. Indeed,
404 if we observe a strong GxE interaction, the ranking of individuals between pots and field will likely vary,
405 hampering an accurate selection of the best individuals. To some extent, this was tested in the across-
406 population scenario and PA values were nearly as high as they were in the within-population scenario
407 for some traits for EDGARR population. This hypothesis should be further investigated for more traits
408 and scenarios. For the second hypothesis, if genetic relatedness was already known to affect PA, its
409 magnitude remains unknown. Especially in this study where the VS was composed of inter-specific
410 varieties, while phenotypic data were only available for *Vitis vinifera* varieties. This is the first time that

411 GP has been applied with such different genetic backgrounds between the training and the validation
412 sets. We tested using completely different populations to train the model, and results were
413 encouraging for most traits for EDGARR population, while PA values were smaller in across-population
414 for Martell population.

415 For across-population GP, we showed that LASSO was more often better than GBLUP, compared to the
416 within-population scenario. This observation was also done in another study on grapevine (Brault et
417 al., 2022b).

418 Phenotyping environment

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

422 Potted own-rooted grapevine phenotypes are likely to differ compared to field phenotypes. However,
423 we have not found studies that compared both different varieties and traits related to the harvest.
424 Most studies on pots or greenhouses were focused on disease resistance or drought tolerance. If this
425 kind of breeding is chosen for the future, one should measure the GxE interactions beforehand.

426

427 Grapevine ideotype

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

441 As many other traits could not be included in the ideotype because of the difficulty of phenotyping,
442 one must ensure that the selection intensity is not too high. Thus, enough individuals with genetic

443 diversity must be kept to be phenotyped for costly traits such as wine aromas later in the breeding
444 program.

445 Another solution for grapevine breeding would be to predict the best crosses to realize, based on the
446 cross mean and variance prediction. The proof-of-concept for cross mean was already done in
447 grapevine (Brault et al., 2022b) but it was not applied in a breeding context. Predicting cross variance
448 would allow to select crosses that would result in extreme offspring phenotypes (Neyhart and Smith,
449 2019; Wolfe et al., 2021).

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

455 ## Conclusion

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

461 # Data availability

462 Data and code to reproduce the results are available at: https://doi.org/10.57745/G8PXEJ.

463 Genomic and phenotypic data for half-diallel and diversity panel populations are available at:
464 https://doi.org/10.15454/PNQQUQ.

465 # Contribution statement

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

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.

Conflict of interest

The authors declare that they have no conflict of interest. The authors declare that the experiments comply with the current laws of the country in which they were carried out.

References

Alonso-Blanco, C., Méndez-Vigo, B., 2014. Genetic architecture of naturally occurring quantitative traits in plants: an updated synthesis. Curr. Opin. Plant Biol., Genome Studies and Molecular Genetics 18, 37–43. https://doi.org/10.1016/j.pbi.2014.01.002

Andrade, L.R.B. de, Sousa, M.B. e, Oliveira, E.J., Resende, M.D.V. de, Azevedo, C.F., 2019. Cassava yield traits predicted by genomic selection methods. PLOS ONE 14, e0224920. https://doi.org/10.1371/journal.pone.0224920

Azevedo, C.F., Ferrão, L.F.V., Benevenuto, J., Resende, M.D.V. de, Nascimento, M., Nascimento, A.C.C., Munoz, P., 2023. Using visual scores and categorical data for genomic prediction of complex traits in breeding programs. https://doi.org/10.1101/2023.02.27.530308

Bates, D., Mächler, M., Bolker, B., Walker, S., 2014. Fitting Linear Mixed-Effects Models using lme4. ArXiv14065823 Stat.

Beavis, W.D., Smith, O.S., Grant, D., Fincher, R., 1994. Identification of quantitative trait loci using a small sample of topcrossed and F4 progeny from maize. Crop Sci. USA.

Bécart, V., Lacroix, R., Puech, C., Cortázar-Atauri, I.G. de, 2022. Assessment of changes in Grenache grapevine maturity in a Mediterranean context over the last half-century. OENO One 56, 53–72. https://doi.org/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, https://doi.org/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 GenesGenomesGenetics 11. https://doi.org/10.1093/g3journal/jkab248

Brault, C., Lazerges, J., Doligez, A., Thomas, M., Ecarnot, M., Roumet, P., Bertrand, Y., Berger, G., Pons, T., François, P., Le Cunff, L., This, P., Segura, V., 2022a. Interest of phenomic prediction as an alternative to genomic prediction in grapevine. Plant Methods 18, 108. 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., 2022b. Across-population genomic prediction in grapevine opens up promising prospects for breeding. Hortic. Res. uhac041. https://doi.org/10.1093/hr/uhac041

Browning, B.L., Zhou, Y., Browning, S.R., 2018. A One-Penny Imputed Genome from Next-Generation Reference Panels. Am. J. Hum. Genet. 103, 338–348. https://doi.org/10.1016/j.ajhg.2018.07.015

Consortium, R., Fugeray-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. 0. https://doi.org/10.3389/fgene.2021.629737

Cortázar-Atauri, I.G. de, Duchêne, E., Destrac-Irvine, A., Barbeau, G., Rességuier, L. de, Lacombe, T., Parker, A.K., Saurin, N., Leeuwen, C. van, 2017. Grapevine phenology in France: from past observations to future evolutions in the context of climate change. OENO One 51, 115–126. https://doi.org/10.20870/oeno-one.2017.51.2.1622

Crossa, J., Pérez-Rodríguez, P., Cuevas, J., Montesinos-López, O., Jarquín, D., de Los Campos, G., Burgueño, J., González-Camacho, J.M., Pérez-Elizalde, S., Beyene, Y., Dreisigacker, S., Singh, R., Zhang, X., Gowda, M., Roorkiwal, M., Rutkoski, J., Varshney, R.K., 2017. Genomic Selection in Plant Breeding: Methods, Models, and Perspectives. Trends Plant Sci. 22, 961–975. https://doi.org/10.1016/j.tplants.2017.08.011

Eibach, R., Zyprian, E., Welter, L., Töpfer, R., 2007. The use of molecular markers for pyramiding resistance genes in grapevine breeding 6.

Elshire, R.J., Glaubitz, J.C., Sun, Q., Poland, J.A., Kawamoto, K., Buckler, E.S., Mitchell, S.E., 2011. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. PLOS ONE 6, e19379. https://doi.org/10.1371/journal.pone.0019379

Endelman, J.B., 2011. Ridge Regression and Other Kernels for Genomic Selection with R Package rrBLUP. Plant Genome 4. 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., Bouckenooghe, V., Roques, M., Pinasseau, L., Verbaere, A., Sommerer, N., Cheynier, V., Bacilieri, R., Boursiquot, J.-M., Lacombe, T., Laucou, V., This, P., Péros, J.-P., Doligez, A., 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 GenesGenomesGenetics 12, jkac103. https://doi.org/10.1093/g3journal/jkac103

Fodor, A., Segura, V., Denis, M., Neuenschwander, S., Fournier-Level, A., Chatelet, P., Homa, F.A.A., 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, e110436. 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. https://doi.org/10.18637/jss.v033.i01

Garrick, D.J., Taylor, J.F., Fernando, R.L., 2009. Deregressing estimated breeding values and weighting information for genomic regression analyses. Genet. Sel. Evol. 41, 55. https://doi.org/10.1186/1297-9686-41-55

Heffner, E.L., Sorrells, M.E., Jannink, J.-L., 2009. Genomic Selection for Crop Improvement. Crop Sci. 49, 1. https://doi.org/10.2135/cropsci2008.08.0512

Lush, J.L., 1937. Animal breeding plans. Anim. Breed. Plans.

Mahalanobis, P.C., 1936. On the generalised distance in statistics. Proc. Natl. Inst. Sci. India.

Meuwissen, T., Hayes, B., Goddard, M., 2016. Genomic selection: A paradigm shift in animal breeding. Anim. Front. 6, 6–14. https://doi.org/10.2527/af.2016-0002

Meuwissen, T., Hayes, B., Goddard, M., 2001. Prediction of Total Genetic Value Using Genome-Wide Dense Marker Maps. Genetics 11.

Migicovsky, Z., Sawler, J., Gardner, K.M., Aradhya, M.K., Prins, B.H., Schwaninger, H.R., Bustamante, C.D., Buckler, E.S., Zhong, G.-Y., Brown, P.J., Myles, S., 2017. Patterns of genomic and phenomic diversity in wine and table grapes. Hortic. Res. 4, 17035. https://doi.org/10.1038/hortres.2017.35

Neyhart, J.L., Smith, K.P., 2019. Validating Genomewide Predictions of Genetic Variance in a Contemporary Breeding Program. Crop Sci. 59, 1062–1072. https://doi.org/10.2135/cropsci2018.11.0716

Nicolas, S.D., Péros, J.-P., Lacombe, T., Launay, A., Le Paslier, M.-C., Bérard, A., Mangin, B., Valière, S., Martins, F., Le Cunff, L., Laucou, V., Bacilieri, R., Dereeper, A., Chatelet, P., This, P., Doligez, A., 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. 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, 1383–1389. https://doi.org/10.1093/bioinformatics/btaa981

Parker, A.K., García de Cortázar-Atauri, I., Gény, L., Spring, J.-L., Destrac, A., Schultz, H., Molitor, D., Lacombe, T., Graça, A., Monamy, C., Stoll, M., Storchi, P., Trought, M.C.T., Hofmann, R.W., van Leeuwen, C., 2020. Temperature-based grapevine sugar ripeness modelling for a wide range of Vitis vinifera L. cultivars. Agric. For. Meteorol. 285–286, 107902. https://doi.org/10.1016/j.agrformet.2020.107902

Reynolds, A., TBX, 2015. Grapevine Breeding Programs for the Wine Industry. Elsevier Science.

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, 164. https://doi.org/10.1186/s12870-016-0850-0

Rienth, M., Vigneron, N., Darriet, P., Sweetman, C., Burbidge, C., Bonghi, C., Walker, R.P., Famiani, F., Castellarin, S.D., 2021. Grape Berry Secondary Metabolites and Their Modulation by Abiotic Factors in a Climate Change Scenario–A Review. Front. Plant Sci. 0. https://doi.org/10.3389/fpls.2021.643258

Rocha, J.R. do A.S. de C., Machado, J.C., Carneiro, P.C.S., 2018. Multitrait index based on factor analysis and ideotype-design: proposal and application on elephant grass breeding for bioenergy. GCB Bioenergy 10, 52–60. https://doi.org/10.1111/gcbb.12443

Santos, J.A., Fraga, H., Malheiro, A.C., Moutinho-Pereira, J., Dinis, L.-T., Correia, C., Moriondo, M., Leolini, L., Dibari, C., Costafreda-Aumedes, S., Kartschall, T., Menz, C., Molitor, D., Junk, J., Beyer, M., Schultz, H.R., 2020. A Review of the Potential Climate Change Impacts and Adaptation Options for European Viticulture. Appl. Sci. 10, 3092. https://doi.org/10.3390/app10093092

Schneider, C., Onimus, C., Prado, E., Dumas, V., Wiedemann-Merdinoglu, S., Dorne, M.A., Lacombe, M.C., Piron, M.C., Umar-Faruk, A., Duchêne, E., Mestre, P., Merdinoglu, D., 2019. INRA-ResDur: the French grapevine breeding programme for durable resistance to downy and powdery mildew. Acta Hortic. 207–214. https://doi.org/10.17660/ActaHortic.2019.1248.30

Smith, H.F., 1936. A Discriminant Function for Plant Selection. Ann. Eugen. 7, 240–250. 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., Péros, J.-P., Doligez, A., 2019. A novel high-density grapevine (Vitis vinifera L.) integrated linkage map using GBS in a half-diallel population. Theor. Appl. Genet. 132, 2237–2252. 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.

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, 3947–3960. 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, A.K., de Rességuier, L., Ollat, N., 2019. An Update on the Impact of Climate Change in Viticulture and Potential Adaptations. Agronomy 9, 514. 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., Matus, J.T., Navarro-Payá, D., Orduña, L., Tello-Ruiz, M.K., Vitulo, N., Ware, D., Rustenholz, C., 2023. An improved reference of the grapevine genome reasserts the origin of the PN40024 highly homozygous genotype. G3 GenesGenomesGenetics 13, jkad067. https://doi.org/10.1093/g3journal/jkad067

Viana, A.P., Resende, M.D.V. de, Riaz, S., Walker, M.A., 2016. Genome selection in fruit breeding: application to table grapes. Sci. Agric. 73, 142–149. https://doi.org/10.1590/0103-9016-2014-0323

Werner, C.R., Gaynor, R.C., Gorjanc, G., Hickey, J.M., Kox, T., Abbadi, A., Leckband, G., Snowdon, R.J., Stahl, A., 2020. How Population Structure Impacts Genomic Selection Accuracy in Cross-Validation: Implications for Practical Breeding. Front. Plant Sci. 11, 592977. https://doi.org/10.3389/fpls.2020.592977

Wolfe, M.D., Chan, A.W., 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. https://doi.org/10.1093/genetics/iyab122

Würschum, T., Maurer, H.P., Weissmann, S., Hahn, V., Leiser, W.L., 2017. Accuracy of within- and among-family genomic prediction in triticale. Plant Breed. 136, 230–236. https://doi.org/10.1111/pbr.12465

Xu, S., 2003. Theoretical Basis of the Beavis Effect. Genetics 10. https://doi.org/165: 2259–2268