Title

Genome-wide association and prediction studies using a grapevine diversity panel give insights into the genetic architecture of several traits of interest

Authors

Flutre T[1,4,5], Le Cunff L[2,4], Fodor A[1,4], Launay A[1,4], Romieu C[1,4], Berger G[1,4], Bertrand Y[1,4], Beccavin I[2], Bouckenooghe V[2,4], Roques M[2,4], Pinasseau L[3], Verbaere A[3], Sommerer N[3], Cheynier V[3], Bacilieri R[1,4], Boursiquot JM[1,4], Lacombe T[1,4], Laucou V[1,4], This P[1,4], Péros JP[1,4], Doligez A[1,4]

Addresses

1 AGAP, Univ Montpellier, CIRAD, INRAE, Montpellier SupAgro, Montpellier, France

2 IFV, Le Grau-du-Roi, France

3 SPO, Univ Montpellier, INRAE, Montpellier SupAgro, Montpellier, France

4 UMT Géno-Vigne, Montpellier, France

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

Corresponding author:

Flutre T, +33 (0)1 69 15 61 33, timothee.flutre@inrae.fr

Doligez A, +33 (0)4 99 61 25 03, agnes.doligez@inrae.fr

Summary

To cope with the challenges facing agriculture, speeding-up breeding programs is a worthy endeavor, especially for perennials, but requires to understand the genetic architecture of important traits. To go beyond QTL mapping in bi-parental crosses, we exploited a diverse panel of 279 *Vitis vinifera* L. cultivars. This panel planted in five blocks in the vineyard was phenotyped over several years for 127 traits including yield components, organic acids, aroma precursors, polyphenols, and a water stress indicator. Such an experimental design allowed us to reliably assess the genotypic values for most traits. The panel was genotyped for 60k SNPs by combining an 18K microarray and sequencing (GBS). Marker densification via GBS markedly increased the proportion of genetic variance explained by SNPs, and two multi-SNP models identified QTLs not found by a SNP-by-SNP model. This led to 489 reliable QTLs using the combined microarray-GBS SNPs for 41% more response variables than a SNP-by-SNP model applied to microarray-only SNPs, and many QTLs were new compared to the results from bi-parental crosses. Prediction accuracy ranging from 0.14 to 0.84 for 80% of the response variables was promising for genomic selection, and provided insights into the genetic architecture of each trait when put in perspective with the number of QTLs and heritability.

Keywords

1. grapevine
2. genotyping-by-sequencing
3. genetic architecture
4. GWAS
5. genomic prediction
6. yield components
7. secondary metabolites
8. candidate genes

Introduction

Viticulture currently faces two major challenges, decreasing inputs, especially fungicide treatments, and adapting to climate change, while maintaining berry quality and differentiated wine styles. In this endeavor, both harnessing existing genetic diversity (Wolkovich *et al.*, 2018) and breeding new varieties (Adam-Blondon *et al.*, 2011) are important levers.

For the latter, many studies over the last two decades aimed at deciphering the genetic architecture of traits of interest by QTL mapping (Vezzulli *et al.*, *in press*). However, this approach suffers from several drawbacks: the limited allelic diversity in parents, the low number of recombination events in the progeny, the upward bias of estimated QTL effects, and the under-estimation of the polygenic contribution for prediction purposes (Cardon and Bell, 2001; Xu, 2003). As a result, all traits currently used in marker-assisted selection (Le Cunff, pers. com.; Vezzulli *et al.*, 2019) are controlled by a single or a few major genes, such as resistance to downy mildew and powdery mildew (DiGaspero *et al.*, 2007), black rot (Rex *et al.*, 2014), sex (Marguerit *et al.*, 2009; Picq *et al.*, 2014), berry color (Fournier-Level *et al.*, 2009), seedlessness (Mejia *et al.*, 2011), and Muscat aroma (Duchêne *et al.*, 2009; Battilana *et al.*, 2009).

To overcome these limits, a few genome-wide association studies (GWASs) were performed in grapevine but did not identify many new QTLs, due to various reasons. Myles *et al.* (2011), Zarouri (2016), Migicovsky *et al.* (2017) and Laucou *et al.* (2018) harnessed phenotypic data from genetic resources repositories collected without a proper experimental design. Moreover, the first three cited articles used at most 10k SNPs despite the low extent of linkage disequilibrium (Myles *et al.*, 2011; Nicolas *et al.*, 2016). Zhang *et al.* (2017) focused on a single binary trait with a major QTL, seedlessness. Yang *et al.* (2017) used only 187 SSRs and 96 genotypes. Moreover, most of these studies, as well as Zarouri (2016) which analyzed 36k SNPs in 242 cultivars, and Guo *et al.* (2019) which analyzed 32k SNPs in 179 cultivars, used SNP-by-SNP models to test for association.

However, SNP-by-SNP models do not exploit the potential gain in power of multi-SNP models (Hoggart *et al.* 2008; Zhang *et al.*, 2019). Such models indeed allow to estimate the cumulative contribution of SNPs with small effects (Yang *et al.*, 2010). They can also be extended to more realistic genetic architectures, with both sparse and dense genetic components (Zhou *et al.*, 2013), the former corresponding to the case with few major genes and the latter with many small-effect QTLs. In addition, they provide a natural way to efficiently perform genomic prediction (GP; de los Campos *et al.*, 2013), even for traits with no major QTLs for which marker-assisted selection is not feasible.

Moreover, focusing only on searching for QTLs is prone to criticism (Rockman, 2012). When breeding is a goal, the effects of published QTLs often are overestimated (Xu, 2003) which leads to poor prediction (Meuwissen *et al.*, 2001). When a large panel of genotypes is suitable for genome-wide association studies, it hence is also relevant to use it for genomic prediction.

Consequently, our objective was to perform whole-genome association studies and genomic prediction analyzes for various traits of interest in grapevine breeding, likely to display different genetic architectures. We aimed at finding out to what extent genetic variation contributes to phenotypic variation, how it is organized in sparse and dense genetic components, and how accurate genomic prediction might be before using it adequately for breeding. Our approach builds on a large diversity panel of 279 *Vitis vinifera* L. cultivars (Nicolas *et al.*, 2016) defined from the French collection of genetic resources, overgrafted in the vineyard in five randomized complete blocks. The panel was phenotyped with this experimental design over several years for 127 traits including yield components, organic acids, aroma precursors, polyphenols, and a water stress indicator. The cultivars were genotyped with both microarray and sequencing after a reduction of genomic complexity (genotyping by sequencing, GBS; Barba *et al.*, 2014; Marrano *et al.*, 2017; Klein *et al.*, 2018; Guo *et al.*, 2019), reaching a total of 63k SNPs. QTL detection and genomic prediction were then performed with multi-SNP models assuming different genetic architectures.

4

Material and methods

Plant material and field trial

The panel of 279 cultivars of *Vitis vinifera* L. is weakly structured in three genetic groups, table east (TE), wine east (WE) and wine west (WW), each composed of 93 cultivars (Nicolas *et al.*, 2016). In 2009, at the Domaine du Chapitre of Montpellier SupAgro (Villeneuve-lès-Maguelone, France), the 279 cultivars were over-grafted on 6-year-old vines of cultivar Marselan, itself grafted on rootstock Fercal (C. Clipet, pers. com.), in a complete randomized block design with five blocks (A to E, Fig. S1). Each of the five blocks contained one plant of each panel cultivar as well as a regular mesh of over-grafts of Marselan as control (between 23 and 39 per block). The trial was maintained under the following training system: double cordon and 3300 plants/ha (1 m between plants along the same rank and 2.5 m between plants of successive ranks).

A subset of 23 genotypes of a Syrah x Grenache progeny (2 parents and 21 full-sibs) was also used to assess out-of-sample genomic prediction (Adam-Blondon *et al.*, 2004; Doligez *et al.*, 2013).

Phenotyping

In 2010, 2011 and 2012, three clusters per plant were harvested at maturity, understood here as 20°Brix, hence providing the sampling date (SAMPLDAY, in days since January 1). Were then measured the number of clusters (NBCLU), mean cluster weight (MCW, in g), mean cluster length (MCL, in cm), and cluster compactness (CLUCOMP, on the OIV 204 scale from 1 to 9; OIV, 2009). Among berries from the middle of clusters, one hundred berries were randomly sampled and weighted, providing the mean berry weight (MBW, in g). In 2011-2012 and 2012-2013 winters, the mean cluster width (MCWI, in cm), number of woody shoots (NBWS) and pruning weight (PRUW, in kg) were measured for each plant. In 2011, the veraison date (VER, in days since January 1) was also recorded. Because in 2010 it was the first fruit set after overgrafting and because pruning weight has an effect on phenotypic responses but was not measured in winter 2009-2010, raw phenotypic data from 2010 were visually explored but discarded from further analyses.

Two variables were computed from traits among the ten listed above: the veraison-maturity interval (VERMATU, in days), and plant vigour (VIG) as pruning weight divided by the number of woody shoots per vine (NBWS).

In 2011 and 2012, juices were made from the sampled berries and analyzed to measure $\delta^{13}C$ (D13C) following Gaudillère *et al.* (2001) as detailed in Pinasseau *et al.* (2017a). In 2012 were also measured glucose (GLU), fructose (FRU), malate (MAL), tartrate (TAR), shikimate (SHI) and citrate (CIT), all in $\mu Eq.L^{-1}$, as detailed in Rienth *et al.* (2016).

Six variables were computed from traits among the seven listed above: the sum of glucose and fructose (GLUFRU), glucose divided by fructose (GLUONFRU), malate divided by tartrate (MALTAR), idem for shikimate (SHIKTAR), citrate (CITAR) and the sum of glucose and fructose (GLUFRUTAR).

In 2014 and 2015, the same field trial was used but differently managed, with irrigation applied to blocks C, D and E only (Pinasseau *et al.*, 2017a). As above, three clusters per plant were harvested at 20°Brix, providing the mean cluster weight (MCW, in g). More details on berry sampling and processing, as well as polyphenols and $\delta^{13}C$ measurements and analysis are in Pinasseau *et al.* (2017a), but note that only the cultivars from the panel were phenotyped (i.e., not the control). Moreover, for a given year, all sampled berries from different blocks with the same water treatment were pooled per cultivar. From the available data on the 105 polyphenols in µg per berry (Pinasseau *et al.*, 2017b), a few typos were corrected and the 17 extra variables defined by Pinasseau *et al.* (2017a) were calculated. In addition, two aroma precursors, β-damascenone (BDAM, in $\mu g.L^{-1}$; Kotseridis *et al.*, 1999) and potential dimethyl sulfide (PDMS, in $\mu g.L^{-1}$; Segurel *et al.*, 2005), were also measured. The volume and weight of the juice samples were recorded.

A total of 127 traits were phenotyped, from which 25 extra variables were computed. Because irrigation was applied to some blocks only in 2014-2015, the yield component and water stress indicator data in 2011-2012 and in 2014-2015 were analyzed separately. As a result, a total of 152 response variables were subsequently analyzed.

The sanitary status of cultivars regarding the presence of five viruses (CNa, GLRaV1, GLRaV2, GLRaV3, GFkV) was assessed by ELISA (Clark and Adams, 1977).

Berry weight was phenotyped on the Syrah x Grenache cross in 2011 and 2012 in the same way as on the panel, as detailed in Doligez *et al.* (2013).

Genotyping

Data acquisition and analysis of microarray SNPs

The panel was genotyped as in Laucou *et al.* (2018) with the GrapeReSeq 18k Vitis microarray from Illumina which contains 18047 SNPs. Data processing (Methods S1) resulted in 13,925 SNPs for 277 cultivars. Of these, 11,102 SNPs remained with linkage disequilibrium between SNP pairs below 0.9, and 10,503 SNPs remained with minor allele frequency per SNP above 0.05.

The subset of 23 genotypes from the Syrah x Grenache cross was genotyped on the same microarray.

Data acquisition and analysis of sequencing SNPs

The panel was also genotyped by sequencing (GBS) following Elshire *et al.* (2011). Keygene N.V. owns patents and patent applications protecting its Sequence Based Genotyping technologies. Data processing (Methods S2) resulted in 184,145 SNPs with less than 30% missing data for 283 accessions (the 279 cultivars from the panel as well as three others not used in this study).

Join imputation of microarray and GBS SNPs

Both SNP data sets (13,925 SNPs from the microarray and 184,145 from the GBS) were combined with duplicate removal into a set of SNPs for 277 common cultivars and `197,885` SNPs using coordinates on the 12Xv2 reference sequence (Canaguier *et al.*, 2017). Missing data were imputed with Beagle version 4.1-r862 (Browning and Browning, 2009) with window=1000, overlap=450, ne=10000 and otherwise default parameters. Two final filtering steps were performed, on LD (<= 0.9) resulting in 90,007 SNPs, and on MAF (>= 0.05) resulting in 63,105 SNPs. We also imputed the Syrah x Grenache SNP genotypes similarly using Beagle.

Statistical modeling of phenotypic data

After an exploratory data analysis (Methods S3), each trait was then analyzed using univariate regression models. Given that the number of SNPs was higher than the number of phenotypic observations, and because of the potential presence of genotype-year interactions as well as spatial heterogeneity, the whole analysis was conducted into two phases. In the first phase (this section), estimates of total genotypic values were obtained. In the second phase (next section), these were regressed on SNP genotypes to identify QTLs, estimate their allelic effects and assess prediction accuracy.

For all traits, whether or not spatial correction was applied, a linear mixed model was fitted by maximum likelihood (ML) with all fixed effects from the global model (as detailed in Methods S4) as well as two random effects, for genotype and genotype-year interaction. Because R/MuMIn tests the inclusion of fixed effects only, and not random effects, R/lmerTest version 3 (Kuznetsova *et al.*, 2017) was used. Explanatory variables were kept based on Fisher tests when modeled as fixed, and on likelihood ratio tests when modeled as random, with a threshold on *p* values at 0.05 for both. The final model was then re-fitted by restricted maximum likelihood (ReML) to obtain unbiased estimates of variance components. Assumptions, such as homoscedasticity, normality, temporal and spatial independence, were checked visually by looking at residuals and empirical best linear unbiased predictors (eBLUPs) of genotypic values. Broad-sense heritability (H^2) for phenotypic means (Nanson, 1970) was computed using both the classical formula for balanced designs using the mean number of trials (years) and replicates per trial (blocks), H^2_C, and a generalized estimator for unbalanced designs (Oakey *et al.*, 2006) ignoring

7

genotype-year interactions, H^2_O. Robust confidence intervals for variance components, heritability and genotypic coefficient of variation were obtained by parametric bootstrap as recommended by Schweiger *et al.* (2016), using the percentile method (Carpenter and Bithell, 2009) in the R/lme4 and R/boot packages.

Empirical BLUPs of genotypic values for berry weight were obtained in the same way on the Syrah x Grenache progeny as on the panel.

Statistical modeling of genotypic data

Empirical BLUPs of total genotypic values were regressed on SNP genotypes via univariate models: eBLUP(**g**) = f(M) + **e**, where eBLUP(**g**) is a vector of responses of length N, M is a matrix of P predictors, here SNP genotypes, of dimension NxP, **e** is a vector of errors of length N, and f is a regression function. SNP genotypes can be encoded for additivity (M_a) or dominance (M_d). Only the former is displayed in the following equations, but both additive-only and additive + dominance models were tested. The regression function f encodes the genetic architecture, either sparse in which only a subset of SNPs have a non-zero effect, or dense in which all SNPs have a non-zero effect (Zhou *et al.*, 2013). As the genetic architecture is unknown, several models were tested, differing in the genetic architecture they assume or the algorithms used to fit them.

Genetic architecture assumed sparse

When assuming a sparse architecture, we used two types of models to perform genome-wide association testing and detect QTLs. The first is the SNP-by-SNP model as implemented in GEMMA version 0.97 (Zhou and Stephens, 2012). For each SNP p, eBLUP(**g**) = $\mathbf{1}\,\mu + M_{a,p}\,\beta_p + \mathbf{u} + \mathbf{e}$ where $M_{a,p}$ is a vector with the genotypes at the p^{th} SNP and $\mathbf{e} \sim N_N(\mathbf{0}, \sigma_e^2\,Id)$ with N the Normal distribution of dimension N, **0** a vector of zeros and Id the identity matrix of dimension NxN. Our goal was to test the hypothesis of a null effect of the SNP of interest (β_p=0), while controlling for relatedness between genotypes with a random effect, **u**, having additive genetic relationships as covariance matrix. Controlling the family-wise error rate at 5% to account for multiple testing, the effects of SNPs were deemed significant when the *p* value from the Wald test statistic was lower than the Bonferroni threshold.

The second type of models jointly analyzes all SNPs. Our goal was to select a subset of SNPs with large effects while handling linkage disequilibrium. This predictor selection can be achieved in a frequentist setting *via* stepwise regression (Segura *et al.*, 2012; Bonnafous *et al.*, 2018). This procedure starts with the SNP-by-SNP model, followed by inclusion, at every iteration, of the SNP with the smallest *p* value as an additional fixed effect, until the proportion of variance

8

explained by the polygenic effect is close to zero. The SNP effects deemed significant were those of the best model selected according to the extended BIC (Chen and Chen, 2008). We fitted it with R/ mlmm.gwas v1.0.4 (Bonnafous *et al.*, 2018) allowing a maximum of 50 iterations. Predictor selection can also be achieved in a Bayesian setting via the variable selection regression model (BVSR): eBLUP(**g**) = **1** μ + M_a **β** + **e**, with the so-called spike-and-slab prior, $\beta_p \sim \pi_0\ \delta_0 + (1 - \pi_0)$ $N(0, \sigma_\beta^2)$, where δ_0 is a point mass at zero. We fitted it with a Bayesian variational algorithm as implemented in R/varbvs version 2.5.7 (Carbonetto and Stephens, 2012). Compared to stepwise frequentist models, varbvs provides point estimates and uncertainty intervals of the proportion of SNPs with a non-zero effect, π_0, as well as of the "SNP heritability" (Yang *et al.*, 2010). Moreover, compared to the same Bayesian model fitted with MCMC as implemented for example in R/BGLR (Perez and Gustavo, 2014), varbvs can be faster by several orders of magnitude, especially with large numbers of predictors. SNPs were deemed significant when their posterior inclusion probability, $PIP_p = Pr(\beta_p \neq 0)$, was larger than 0.80.

QTL definition and annotation

QTLs were defined as intervals around significant SNPs based on the decay of linkage disequilibrium similarly to Bonnafous *et al.* (2018), as detailed in Methods S5. They were annotated using the genomic annotations from Canaguier *et al.* (2017). We also used the correspondence between IGGP (International Grapevine Genome Program) and NCBI RefSeq gene model identifiers provided by the URGI (https://urgi.versailles.inra.fr/Species/Vitis/Annotations). A comparison was performed between the QTLs detected in this study and a list of already-published QTLs (Vezzulli *et al.*, *in press*; QTLs significant at a 5% genome-wide threshold) that were classified according to the Vitis INRAE ontology v2 (Duchêne, 2020) and slightly edited for automatic processing. This comparison was made only at the chromosome level because genomic coordinates on the reference genome were difficult to retrieve from publications, and sometimes impossible especially when other Vitis species and interspecific hybrids were involved. A similar comparison was performed with significant hits from a few GWAS publications after converting their coordinates on the genome reference we used.

Genetic architecture assumed dense

When assuming a dense architecture, the multi-SNP model is the ridge regression: eBLUP(**g**) = **1** μ + M_a **β** + **e** where **β** ~ N_P(**0**, σ_β^2 Id). Our first goal was to estimate the proportion of variance of empirical BLUPs of genotypic values explained by SNPs (PVE_{SNPs}) to assess the need for additional SNPs. The classical parameterization of genotypic values in additive values and dominance

9

deviations was used with the appropriate design and covariance matrices based on SNP genotypes (VanRaden 2008, Vitezica *et al.*, 2013) so that there is an equivalence between the classical "animal model" and the ridge regression (Habier *et al.*, 2007): eBLUP($\mathbf{g}$) = $\mathbf{1}\,\mu + \mathbf{g_a} + \mathbf{e}$ where $\mathbf{g_a} \sim N_N(\mathbf{0}, \sigma_a^2 A)$ where A, the NxN matrix of additive genetic relationships, is proportional to the matrix product $M_a M_a^T$ once M_a is centered using allele frequencies. We implemented this model in R/lme4 version 1.1.19 (Bates *et al.*, 2015) and computed confidence intervals for variance components by bootstrap as above. When the variance component for dominance deviations was included, the algorithm often did not converge. Because the estimators of additive and dominance relationships from SNPs assume linkage equilibrium, a threshold on LD of 0.5 was applied.

Genomic prediction

The multi-SNP models, whether assuming a sparse or dense genetic architecture, also estimate SNP effects allowing out-of-sample prediction (Meuwissen *et al.*, 2001). This was assessed within the panel by K-fold cross-validation, with K set at 5 (Arlot and Lerasle, 2016), repeated 10 times, with R/caret version 6 (Kuhn, 2018), using R/varbvs for the sparse architecture and R/rrBLUP version 4.5 (Endelman, 2011) for the dense architecture. We assessed prediction accuracy between empirical BLUPs of genotypic values and their predictions with a range of metrics: root mean square error (RMSE); Pearson's linear correlation coefficient (corP) and Spearman's rank correlation coefficient (corS); as well as outputs from the simple linear regression of observations on predictions (Pineiro *et al.*, 2008) such as the intercept, slope, adjusted coefficient of determination (R^2) and *p* value of the test for no bias (Baey, 2014).

Out-of-sample prediction was also assessed by training rrBLUP and varbvs methods on the whole panel and predicting empirical BLUPs of genotypic values from the 23 genotypes of the Syreah x Grenache cross.

Reproducibility

Given the amount of resources needed to perform a genome-wide association study with a proper experimental design in a perennial plant species, we chose to implement our analyzes in such a way that it allows methods reproducibility in the sense of Goodman *et al.* (2016). Demultiplexed reads were inserted into the SRA database of the NCBI as BioProject PRJNA489354. We also made available other data and computer code on data.inrae.fr (if not specified otherwise), as detailed in Methods S3.

10

Results

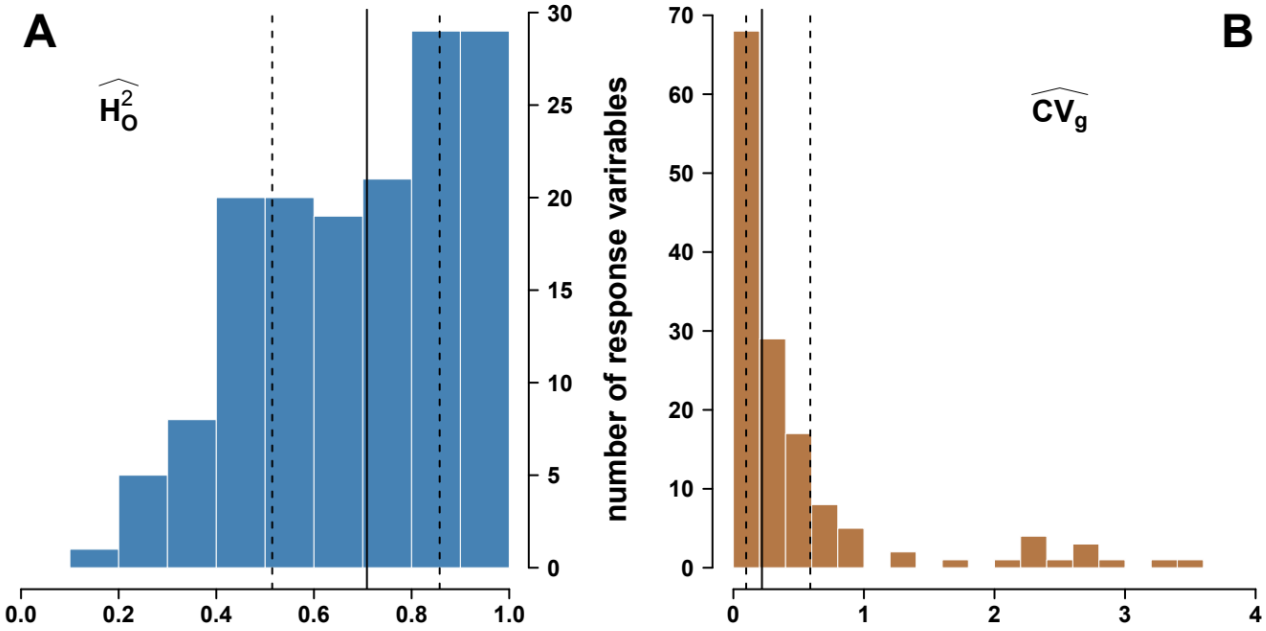
Estimation of broad-sense heritabilities and genetic coefficients of variation

All 152 analyzed response variables displayed substantial variation after conditioning on year, block and irrigation (Fig. S2). For some polyphenol variables, part of the variation was obviously associated with skin color (not shown). For the 25 response variables with data in 2011 or 2012, thanks to the control, we could assess that part of this variation is of genetic origin. For mean berry weight, a narrow distribution of control data suggested a large part of genetic variation, but a visual inspection shows that this was not the case for the other response variables (Fig. S2). We looked for spatial heterogeneity using the control regularly planted in each block. As variograms were mostly flat (Fig. S3) and prediction errors assessed by cross-validation were high (not shown), we concluded that spatial correction was not necessary. Depending on the response variable, the amount of missing data ranged from 15.78% to 43.93% (Table S2). To account for such unbalance when controlling for known confounders, we fitted linear mixed models and obtained the BLUPs of the genotypic values. After model selection, the final set of fixed and random effects differed between response variables (Table S2).

As shown in Figure 1, 76.6% of the broad-sense heritabilities (H^2) were above 0.5 (arbitrarily chosen here as a quality threshold), with narrow confidence intervals (Table S2). Two different estimators, H^2_C and H^2_O, handling missing data differently, gave very similar estimates (Table S2). This measure of experiment accuracy indicated that, for most response variables, the phenotypic data of a given cultivar provided a high degree of agreement with the genotypic value of this cultivar. Moreover, 92.7% of the genetic coefficients of variation are above 5% and 59.1% above 20% (Figure 1, Table S2).

Fig. 1 Estimation in a diverse panel of *Vitis vinifera* L. of (A) broad-sense heritabilities for 152 response variables using the estimator from Oakey *et al.* (2006), H^2_O, and (B) their genetic coefficients of variation, CV_g. Vertical lines indicate the median (plain), and quantiles at 0.25 and 0.75 (dotted).

11



Combining genotyping technologies to explain more genetic variance

Once we obtained the genotypic BLUPs of all cultivars for each response variable, we aimed at explaining their variance with SNP genotypes. For that purpose, we used two sets of SNPs, hereafter referred to as “microarray-only SNPs” and “microarray-GBS SNPs”, obtained as follows.

Nicolas *et al.* (2016) originally defined the population membership of each cultivar with 20 SSRs using STRUCTURE (Pritchard *et al.* 2000, Falush *et al.* 2003). We did here a DAPC using 8840 microarray SNPs without any missing data. This confirmed the genetic structure in three weakly differentiated clusters, called “population” hereafter. When performing a PCA, the first principal component accounted for 8.1% of the total variance, and the second one for 2.8% (Fig. S4). Moreover, results from SNPs revealed a change in population membership for nine cultivars (Fig. S4 and Table S3), most probably due to a better genome coverage. Most SNPs had moderate allele frequencies, and cultivars from the Wine West population had a deficit of low-frequency SNPs (Fig. S5), in agreement with the ascertainment bias typical of microarray-based high-throughput genotyping (Albrechtsen *et al.*, 2010). Only cultivars from the Table East population showed a slight excess of low-frequency SNPs. After filtering on LD below 0.9 and MAF below 0.5, 10,503 SNPs remained, which formed the first set of SNPs (“microarray-only SNPs”) to be used in GWAS and genomic prediction.

Because LD is known to be short in *Vitis vinifera* L. (Myles *et al.*, 2011; Nicolas *et al.*, 2016), we increased the SNP density by sequencing with complexity reduction (GBS) using the *Ape*KI restriction enzyme. Raw reads had high quality along their sequences, although many displayed adapter content at their 5’ end, which had to be trimmed off. After demultiplexing, more

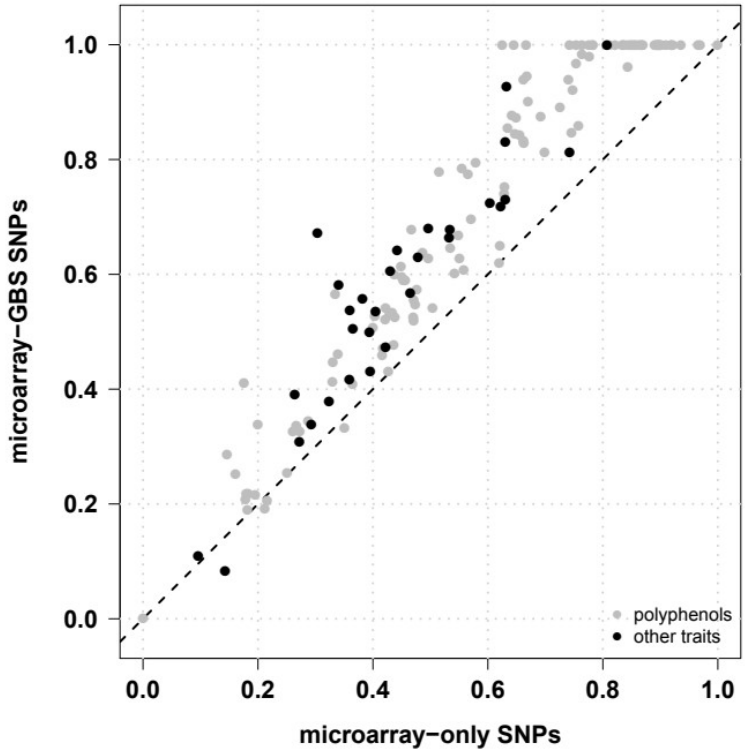
12

than 95% of the reads were assigned to a cultivar. After alignment on the reference genome, the depth of coverage (Molnar and Ilie, 2015) of regions having at least one read, averaged over cultivars, ranged from a minimum of 2.3 reads to a maximum of 81.2 with a median at 21.7. This indicated a reasonable chance of properly calling both homozygous and heterozygous SNPs. After filtering out SNPs with calling quality below 20 and supported by less than 10 reads, and setting as missing SNP genotypes with more than 30% missing data, 184,145 SNPs remained.

We combined microarray SNPs with GBS SNPs, reaching a total of 197,885 SNPs for 277 cultivars in common. Missing data were imputed using LD with Beagle as advised by Swarts *et al.* (2014) for highly heterozygous samples with unknown segregating parental haplotypes. After filtering SNPs on LD above 0.9, 90,007 SNPs remained. The distributions of allele frequencies were similar in the three populations (Fig. S5). Moreover, as expected from sequencing compared to microarrays, they showed an excess of low-frequency SNPs. After filtering on MAF below 0.05, we used the combined data set of 63,105 SNPs ("microarray-GBS SNPs") for GWAS and genomic prediction.

Most importantly, compared to the microarray-only SNP set, the combined microarray-GBS set displayed a substantially higher SNP density along all chromosomes (Fig. S6). We hence computed the proportion of variance in genotypic BLUPs explained by SNPs (PVE_{SNPs}). For this, we estimated the genetic relationships between cultivars (Fig. S7). When assuming an additive-only, polygenic architecture, for the vast majority of responses variables (97.8%), PVE_{SNPs} was higher with microarray-GBS SNPs than with microarray-only SNPs (Fig. 2, Table S4). This clearly showed the advantage of combining SNPs to increase the likelihood that the QTLs are in LD with at least one genotyped SNP.

Fig. 2 Estimation in a diverse panel of *Vitis vinifera* L. of the proportion of variance in genotypic BLUPs explained by SNPs for 152 response variables and two SNP densities, assuming an additive-only, polygenic architecture.



Models including both additive and dominance relationships converged with difficulty. Morever, the proportion of variance of genotypic BLUPs explained by microarray-GBS SNPs when both additive and dominance relationships were included was always equal or lower than with only additive relationships (Table S4). The matrix of dominance relationships was very similar to the identity matrix, making it virtually indistinguishable from the error term (Fig. S7). The genetic variance component of dominance and the error variance hence were unidentifiable.

QTL detection by GWAS and identification of candidate genes

First, each of the 152 response variable was separately analyzed with a SNP-by-SNP model fitted using GEMMA. With the microarray-only SNPs, we detected a total of 2,295 significant SNPs for 88 response variables and, with the microarray-GBS SNPs, 7,855 significant SNPs for 101 response variables (Table 1 and Table S5). For each response variable, because SNPs can be in LD with each other, we defined an interval around each significant SNP using the 95% quantile of kinship-corrected LD between random SNP pairs and the distance in bp predicted for this threshold. In the following, each such interval is called a QTL. Using the microarray-GBS data set, 2.8 million SNP pairs gave a LD threshold of 0.056 corresponding to a 50-kb distance (Fig. S8). The QTL around each significant SNP hence consisted in a physical interval of 100 kb. After merging the

14

overlapping QTLs per response variable, the SNP-by-SNP model identified a total of 1,179 QTLs with the microarray-only SNPs and 1,784 QTLs with the microarray-GBS SNPs (Tables 1 and S6).

Then, to benefit from a potential gain in power when detecting significant SNPs and accuracy when estimating their effects, we fitted two multi-SNP models, using mlmm.gwas and varbvs. With the microarray-only SNPs, mlmm.gwas detected a total of 1,257 significant SNPs corresponding to 1,243 QTLs for 148 response variables and, with the microarray-GBS SNPs, 703 significant SNPs corresponding to 692 QTLs for 125 response variables (Tables 1, S5 and S6). With the microarray-only SNPs, varbvs detected a total of 266 significant SNPs corresponding to 257 QTLs for 118 response variables and, with the microarray-GBS SNPs, 258 significant SNPs corresponding to 257 QTLs for 119 response variables (Tables 1, S5 and S6).

Table 1 Comparison between methods in terms of the number of QTLs (#QTLs) identified in a diverse panel of *Vitis vinifera* L. for two SNP data sets, summed up over all response variables. Also indicated are the number of response variables with at least one QTL (#RVs), and the number of significant SNPs (#sSNPs).

Method		**microarray-only SNPs**	**microarray-GBS SNPs**
Model	**Software**	**#RVs: #sSNPs ; #QTLs**	**#RVs: #sSNPs ; #QTLs**
SNP-by-SNP	GEMMA	88: 2,295 ; 1,179	101: 7,855 ; 1,784
multi-SNP	mlmm.gwas	148 : 1,257 ; 1,243	125: 703 ; 692
	varbvs	118: 266 ; 257	119: 258 ; 257

For both SNP data sets, the number of response variables with at least one QTL was higher with the multi-SNP methods than with the SNP-by-SNP method, confirming the gain in power obtained with multi-SNP models. Within multi-SNP methods, mlmm.gwas found more significant SNPs and QTLs than varbvs, and for more response variables. Yet, the interpretation is not straightforward as, notably, these methods do not use the same criterion for declaring a SNP as significant (see Discussion). Surprisingly, for both multi-SNP methods, the number of response variables with at least one QTL was lower with more tested SNPs, as well as the numbers of significant SNPs and QTLs.

We merged all QTLs per response variable over both SNP sets and all three methods. This yielded a total of 3,490 QTLs over 150 response variables (Table S7), which corresponded to an increase of 196% in the number of QTLs and of 70% in the number of response variables with at least one QTL, compared to applying the SNP-by-SNP method on the microarray-only SNPs. Over

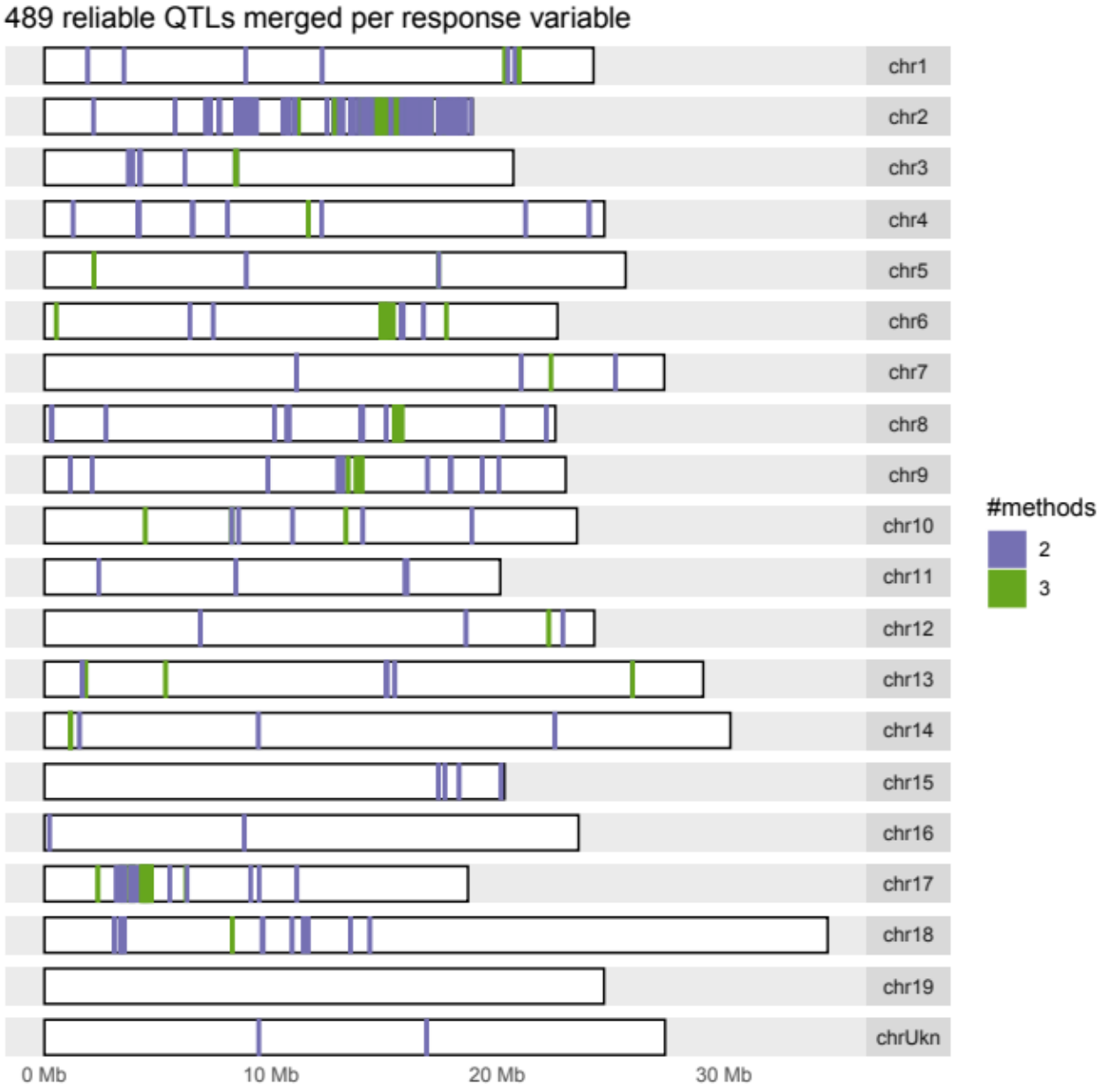
the 3,490 QTLs, 136 were found by all three methods, while 3,001 were found by a single method only and 1,598 by multi-SNP methods only (Fig. S9). Response variables with at least one QTL had a median number of QTLs of 23 and a maximum of 68. Furthermore, over these 150 response variables, 26 had no QTL according to the SNP-by-SNP method but at least one found by both multi-SNP methods (Fig. S10).

In terms of genomic distribution, all chromosomes harbored at least one QTL (Fig. S11), and most QTLs found only by the multi-SNP mlmm.gwas method fell far from QTLs found by other methods (Fig. S12). Moreover, 90% QTLs found only by the SNP-by-SNP method GEMMA clustered on chromosome 2 for 80 response variables (all of them but three being polyphenols, in relation with the anthocyanin-related MYB genes on this chromosome, Matus *et al.*, 2008). This illustrates the fact that such a method reports all significant SNPs whatever the LD pattern between them (Fig. S12). In contrast, the multi-SNP varbvs method was more parsimonious, yet had enough power to identify significant SNPs in regions in which GEMMA did not identify any signal.

In an attempt to identify a reduced set of QTLs with high priority for further investigation, 489 QTLs involving 124 response variables were deemed the most reliable as they were found by at least two methods (Table S7). They corresponded to 59% less QTLs but 41% more response variables with at least one QTL, compared to applying the SNP-by-SNP method on the microarray-only SNPs. All chromosomes harbored at least one such reliable QTL, except chromosome 19 (Fig. 3). The reliable QTL lengths ranged from 100,001 bp to 1,072,169 bp, with a median at 145,089 bp.

16

Fig. 3 Genomic distribution of the most reliable QTLs identified by two methods in a diverse panel of *Vitis vinifera* L. after merging them over microarray-only and microarray+GBS SNP sets per response variable. The color legend indicates the number of methods that identified a given QTL.



The 489 most reliable QTLs were compared with the largest list of QTLs detected in bi-parental crosses of grapevine compiled so far (Vezzulli *et al.*, *in press*). This list synthesizes information about 535 main QTLs from 78 publications ranging from 2002 to 2019 involving 55 crosses (17 intraspecific, 37 interspecific and one unknown). It concerns a total of 102 traits (more or less specific, e.g., all anthocyanins are grouped together) from seven classes specified as in the *Vitis* INRAE ontology. Among the 149 traits analyzed in our study, 128 were deemed absent from

the list of published QTLs, for which we found 448 reliable QTLs, and 21 deemed present, accounting for the 41 other reliable QTLs, as listed in Table S8. For these 21 traits in common, QTLs on the same chromosome were found only for six traits (Table S7): cluster number (on chromosome 7), berry weight (on chromosomes 1, 2, 8, 11, 15 and 17), malate (on chromosomes 9 and 18), and (un)methylated anthocyanins (on chromosome 2), glucose to fructose ratio (on chromosome 2). Therefore, when summing up at the QTL level over all response variables, among our 489 reliable QTLs, only 4.7% were on the same chromosome as published main QTLs.

We also compared our reliable QTLs with significant GWAS hits from other publications in grapevine. Only two traits (cluster and berry weights) were phenotyped in at least one other study and for which at least one significant GWAS hit was found (Zarouri, 2016; Laucou *et al.*, 2018; Guo *et al.*, 2019). For berry weight, out of the 10 QTLs we found, 8 were deemed new on chromosomes 1, 2, 8, 11, 15 and 17. We also found two QTLs on chromosome 8 close to a GWAS hits from Zarouri (2016), but did not recover other GWAS hits from Zarouri (2016) on chromosomes 5 and 17, and from Guo *et al.* (2019) on chromosomes 17, 18 and 19. For cluster weight, we found two new QTLs on chromosomes 1 and 3 but did not recover the GWAS hits from the other studies, on chromosomes 5 (Zarouri, 2016) and 13 (Laucou *et al.*, 2018).

A drawback of QTL detection is its focus on statistical significance, a dichotomization of evidence known to have several limitations (McShane and Gal, 2018). It is usually recommended to, at least, also check and provide effect estimations (Gardner and Altman, 1986). All estimates of significant additive SNP effects are hence given in supplementary (Table S5), along with a quantification of their uncertainty. For each of the 489 reliable QTLs, we also provide a boxplot per genotypic class for one of the significant SNP, arbitrarily chosen among those associated with the QTL (Fig. S13).

To help highlighting candidate genes, we compared the reliable QTLs with the reference genomic annotations gathering 42,413 gene models. As the same locus can be a QTL for multiple response variables, we first merged the 489 QTLs across all response variables, which resulted in 134 distinct genomic intervals (Table S9). These intervals had a median length of 100,001 kb (with a minimum of 100,001 kb and a maximum of 1,072,169 kb). The comparison with gene models yielded 1928 hits with 1926 distinct gene models (Table S10). The median number of overlaps per interval was 11, with a minimum of 2 and a maximum of 87. Among the 1926 gene models, 1313 had a NCBI RefSeq identifiers. Out of these, 333 where annotated as “uncharacterized locus” and hence 980 had an annotations among 863 distinct ones (Table S11).

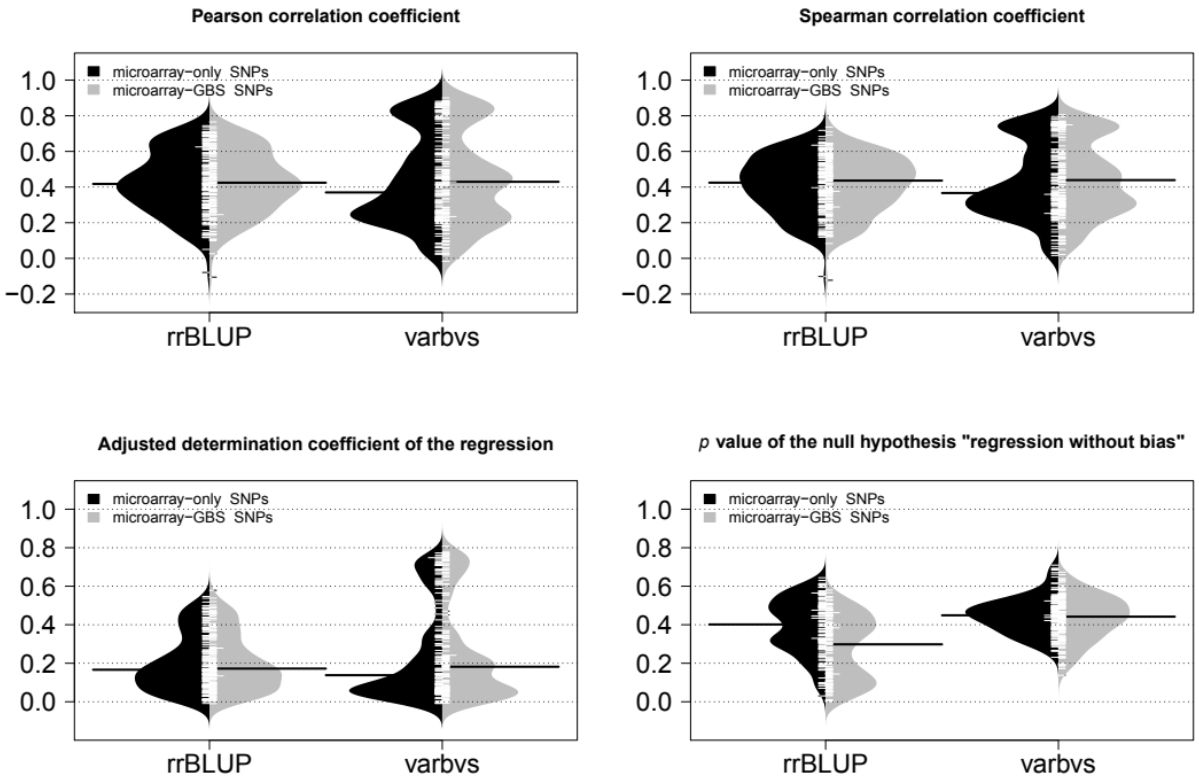
18

As shown on Fig. S11 and S12, a large portion of chromosome 2 (between 12 Mb until the end at 18 Mb) displays a high density of QTLs due to the large number of response variables linked to polyphenols.

Assessment of genomic prediction and insight into genetic architectures

As a first step, we assessed the accuracy of genomic prediction within the panel of 279 cultivars, using repeated *K*-fold cross-validation. Two methods were compared, the first one assuming a sparse genetic architecture, with R/varbvs as its GWAS results (above) showed how parsimonious yet powerful it was, and the second one assuming a dense genetic architecture, with R/rrBLUP implementing the ridge regression corresponding to the infinitesimal model as a baseline. Note that the QTL results from the GWAS section were not used when training each model, to avoid overfitting. Then, for each test set of the cross-validation, various metrics were computed to compare the genotypic BLUPs obtained from phenotypic data only and the predictions obtained from additive SNP effects only (Table S12).

Fig. 4 Assessment of genomic prediction accuracy within a diverse panel of *Vitis vinifera* L. with microarray-only and microarray-GBS SNPs for 152 responses variables by repeated *K*-fold cross-validations. The four metrics were averaged over folds and replicates.



19

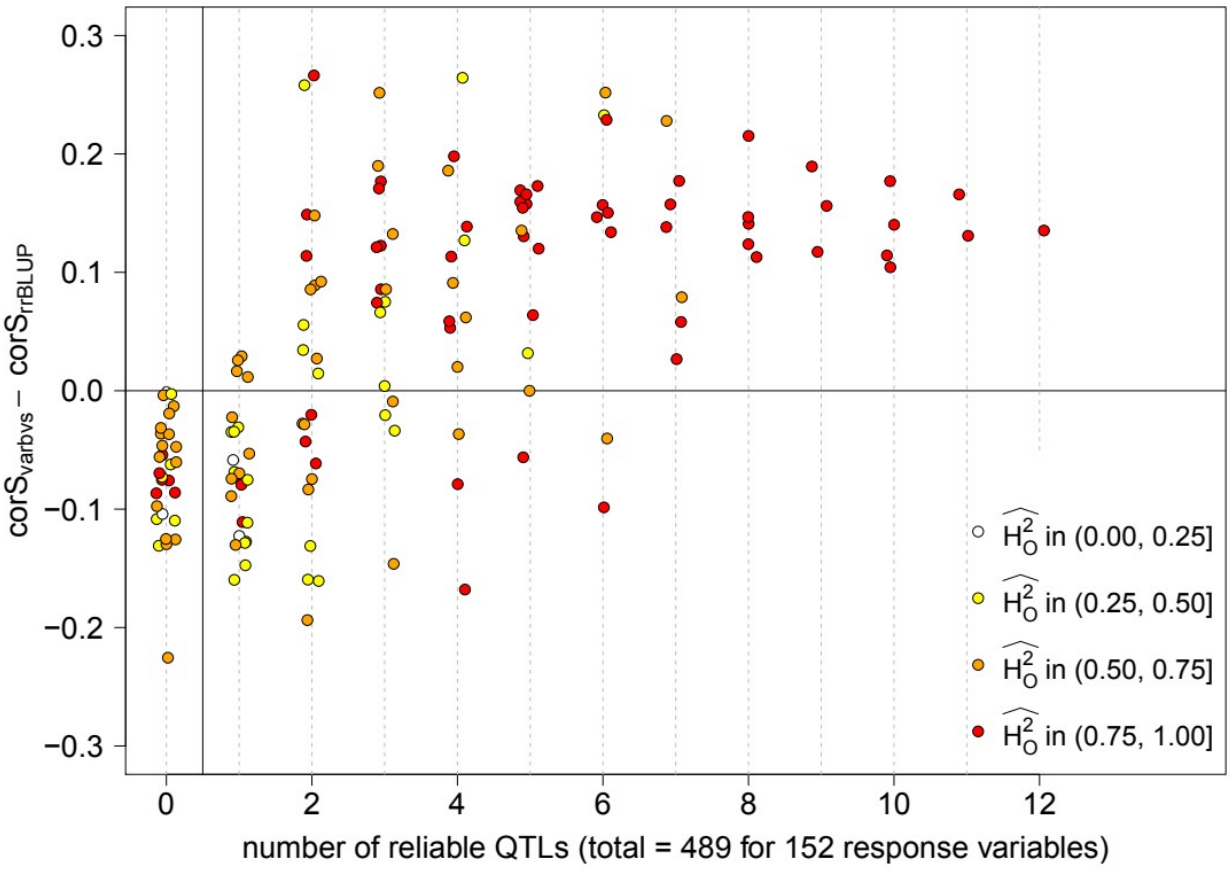
As shown in figure 4, the median Pearson and Spearman correlation coefficients fell between 0.37 and 0.44, with 80% of the whole distributions ranging between 0.14 and 0.84. Comparisons between these correlation coefficients and the broad-sense heritability of each response variable showed a substantial correlation (Fig. S14), higher for varbvs (~0.65) than for rrBLUP (~0.54). For both methods, most Spearman coefficients from the genomic prediction are lower than the broad-sense heritabilities. This was expected since the genomic prediction models we tested only exploited additive genetic variance. But note that low values of Spearman coefficients (below 0.2), and a few negative ones, occurred for traits with medium broad-sense heritabilities (between 0.4 and 0.7). Based on figure 4, both methods had similar median correlation coefficients. However, the distributions of rrBLUP's correlation coefficients were roughly uni-modal whereas varbvs' clearly were multi-modal. This confirmed what was known from simulations (e.g., Wang *et al.*, 2015), that rrBLUP's assumption of an infinitesimal architecture is fairly robust compared to varbvs' assumption of a sparse architecture, yet varbvs can provide substantially better predictions than rrBLUP for some traits. Moreover, rrBLUP results did not seem to depend on the SNP set whereas, for varbvs, results were slightly better with the microarray-GBS SNPs. This suggests that, among the extra SNPs provided by GBS, varbvs managed to identify those which improved its predictions. When looking at the determination coefficient, the median for rrBLUP (0.17) also did not depend on the SNP sets and both distributions looked fairly similar. In contrast, the median for varbvs increased from 0.14 with microarray-only SNPs to 0.18 with microarray-GBS SNPs. The 0.80 quantile for rrBLUP was around 0.44 whereas for varbvs it was around 0.70. Moreover, concerning the *p* value of the test for no bias, varbvs showed similar values across both SNP sets, higher than rrBLUP in general and above 0.05, suggesting an absence of bias. On the contrary, rrBLUP with the microarray-GBS SNPs showed lower *p* values compare to with the microarray-only SNPs. This suggests that the constraint from the infinitesimal model behind rrBLUP to estimate all SNP effects to be non-zero may be too far from the real genetic architecture, especially when SNP density is high.

As a second step, we assessed the accuracy of genomic prediction using the panel of 279 cultivars as a training set to predict mean berry weight in a subset of a Syrah x Grenache progeny. With rrBLUP (respectively varbvs), this gave a Pearson correlation of 0.56 (0.35) and Spearman correlation of 0.54 (0.26), an adjusted coefficient of regression of 0.28 (0.08), and a *p* value when testing for no bias of 1.6×10^{-4} (3.5×10^{-3}). These values are promising, even though the adjusted coefficient of regression is rather weak, and predictions are biased. Moreover, rrBLUP gave better correlations than varbvs, which was in agreement with the results obtained by cross-validation within the panel (Pearson correlation of 0.71 with rrBLUP and 0.61 with varbvs).

20

Finally, combining results from both QTL and genomic prediction can provide insight into the genetic architecture of the studied traits. As shown in Figure 5 (made from data in TableS13), the more reliable QTLs a response variable had, the more accurately varbvs predicted the BLUPs of its genotypic values compared to rrBLUP, which would suggest that these traits have a sparse architecture. In contrast, rrBLUP predicted better than varbvs the response variables for which less than 6 QTLs were detected, and notably the case where at most 1 QTL was found, suggesting here a dense architecture for these traits. Yet, coloring points with respect to broad-sense heritability shows that response variables for which varbvs predicted better than rrBLUP seemed to have not only more reliable QTLs but also a higher broad-sense heritability.

Fig. 5 Interplay between the number of QTLs deemed reliable, the difference in prediction accuracy between methods, and broad-sense heritability, using 152 response variables phenotyped on a diverse panel of *Vitis vinifera* L. Prediction accuracy corresponds to the Spearman correlation coefficient averaged over cross-validation folds and replicates when using the microarray-GBS SNP set. Broad-sense heritability was estimated based on Oakey *et al.* (2006).



21

Discussion

For most traits, high genetic coefficient of variation (CV_g) indicated a substantial amount of genetic variation around the mean value, which suggested promising opportunities for selection. It hence motivated the detection of QTLs and the estimation of their effects, as done in GWAS, and the prediction of breeding values, as done in genomic selection, which are two sides of the same coin. Indeed, both gain from deciphering the genetic architecture of traits of interest. In this challenge, three key components are interlinked, phenotypic data, genotyping data and statistical models, all three of which requiring us to choose between alternatives with trade-offs. We discuss ours in the following and suggest avenues of improvement, of interest to perennial crops in general and grapevine in particular.

Design and analysis of the field trial

Acquiring phenotypic data from which genotypic values can be deduced with sufficient accuracy is a big challenge, especially because a large panel is a prerequisite to have enough power to detect QTLs (Nicolas *et al.*, 2016). Our randomized block design certainly helped in reaching high broad-sense heritabilities for certain traits, yet others show lower ones (see also the sometimes large variation among controls in Fig. S2). Some classical, by-hand phenotyping procedures, when performed on a large panel in the field, are very time-consuming, requiring the coordination of enough manpower in an error-prone process. This calls for the implementation, testing and deployment of high-throughput methods in complement or replacement (Fiorani and Schurr, 2013; Kicherer *et al.*, 2017). But different strategies need to be assessed, notably in terms of investment (Reynolds *et al.*, 2019). Another, major challenge consists in sampling items, such as fruits, at a similar physiological stage, otherwise leading to unknown confounders impossible to control within the statistical model. This is a particularly pressing issue for grapevine due to the strong intra- and inter-cluster heterogeneity between berries (Shahood, 2017). New protocols were proposed, requiring temporal sampling, but work remains to be done to automatize them allowing the phenotyping of a large number of genotypes (Bigard *et al.*, 2018).

In terms of statistical modeling, we chose a two-stage procedure for ease of analysis (Möhring and Piepho, 2009). To comply with the assumptions of the linear mixed model used in the first stage, we had to transform the raw phenotypic data for several traits based on visual assessment. An alternative could have been to apply a more statistically-motivated transformation (Box and Cox, 1964; Burbidge *et al.*, 1988), but these ones apply only to linear models, *i.e.*, without random effects. An avenue of improvement would be to try extensions of the Box-Cox family of

transformation to linear mixed model (Gurka *et al.*, 2006), and assess how well they perform model selection. Another, major decision was what to include as explanatory factors in the full model before model selection. We chose to include pruning weight, the number of wooding shoots and vigour, but neither flower sex nor berry color. Our rationale was that the former three are mainly influenced by location as well as the way the field trial is conducted, whereas the latter two are fully determined genetically, even though flower sex can be converted by manipulation (Negi and Olmo, 1966). We assumed that excluding those strongly genetically-determined from the explanatory factors at the first stage of the analysis would allow to keep most genetically-based variation between genotypes for the second stage of the analysis (GWAS and genomic prediction). Another direction for future work would be to exploit the correlations between traits by using multivariate models (Mardia *et al.*, 1979). Indeed, Pearson correlation coefficients between the BLUPs of response variables showed some patterns (Table S14 and Fig. S15). A comparison of univariate and multivariate linear models could be done at the first stage of the analysis, and SNP-by-SNP versus multi-SNP multivariate models could also be performed at the second stage, both comparisons being the subject of a future article. A more ambitious approach would be to analyze several traits jointly guided by process-based models such as functional-structural plant models (Sievanen *et al.*, 2014), be they at the organ or plant level (Génard *et al.*, 2010; Pallas *et al.*, 2009). This would allow the investigation of genotype-environment interactions, but would notably require the phenotyping of all key phenological stages.

Increase of genotyping density

When genotyping a sample to perform a GWAS, one aims at having a marker density so that each causal locus has a high probability of being in strong enough LD with at least one marker (Kruglyak, 1999). The specific number of required markers depends on the evolutionary process of the sample under study, but in grapevine half a million SNPs may be the minimum (Nicolas *et al.*, 2016; Myles *et al.*, 2009). Reaching such numbers would require whole-genome sequencing. The cost of fully sequencing this panel of 279 genotypes may still be too high for some time. In addition, even though the sequencing techniques keep improving (Jung *et al.*, 2019), highly heterozygous genomes require the complex assembly of genomic fragments. As an intermediate step, genotyping by sequencing the same genotypes as we did here but with another restriction enzyme could increase the final SNP density as long as sequenced locus are different enough between enzymes, which can be explored *in silico* (https://github.com/timflutre/insilicut).

Imputation of heterozygous genotypes from GBS data such as ours is notoriously difficult (Swarts *et al.*, 2014). Moreover the large amount of missing data makes it difficult to

23

626 properly assess imputation accuracy similarly to the cases where dense reference haplotypes are
627 available (Marchini and Howie, 2010). To validate our microarray-GBS set in a way that is linked
628 to our main interest, the association between genotypic and phenotypic data, we looked at the
629 proportion of variance in BLUPs of genotypic values explained by SNP genotypes (PVE_{SNPs}). The
630 improvement when going from the microarray-only set to the microarray-GBS set increased our
631 degree of trust in the genotyping and imputation procedures. Yet, PVE_{SNPs} did not equal 1 for all
632 response variables. Many factors can underly this discrepancy. First, empirical BLUPs of genotypic
633 values are not fully accurate versions of the "true" genotypic values, as reflected in the distribution
634 of broad-sense heritabilities already discussed above. Second, the microarray-GBS set may not tag
635 the core genome of the panel well enough, with a SNP density being too low and pan-genome
636 structural variations remaining undetected, an issue which would be fixed by whole-genome
637 sequencing (Marroni *et al.*, 2014). Third, the assumptions of our linear mixed model may be unmet
638 in the data. Even if the additive relationships we included are supposed to capture the effect of
639 genetic structure (Astle and Balding, 2009) and that models we tested including dominance
640 relationships did not converge, alternative models could be tested, notably those robust to outliers
641 (Gianola *et al.*, 2018) or those capturing nonlinear allelic effects (Jacquin *et al.*, 2016).
642
643 **Sensitivity and specificity of QTL detection, and candidate genes**
644 We have endeavored to compare three methods of genome-wide association studies, using them as
645 most practitioners do in practice. But unfortunately such a comparison effort quickly reaches its
646 limits. Indeed, most practitioners use such methods in a hypothesis testing context to identify a set
647 of significant SNPs, hence dichotomizing evidence in the data. Because the methods minimize
648 different criteria (family-wise error rate, false discover rate) and handle the multiple testing issue in
649 different ways (SNP-by-SNP testing followed by a *p* values correction, or joint multi-SNP
650 selection), a SNP can be declared significant by one method and not by another, even though it is
651 slightly above the threshold of the former and slightly below the threshold of the latter. Another,
652 major misleading factor when comparing GWAS methods is linkage disequilibrium. Comparing
653 SNP-by-SNP and multi-SNP methods in terms of the total number of significant SNPs is not as
654 relevant as it seems as SNP-by-SNP methods do not take LD into account. Moreover, two different
655 multi-SNP methods can select two different, yet linked SNPs for arbitrary reasons, such as the
656 initial order of these SNPs as given to the software implementing the method. That is why the
657 number of significant SNPs reported per method varies widely. The very high number from the
658 SNP-by-SNP method does notably not indicate a better power compare to the multi-SNP methods.
659 When performing a GWAS, it helps keeping in mind that, given the dimension of the data set (*n*

genotypes and *p* SNPs), hypothesis testing becomes hopeless when the number, *k*, of truly associated SNPs is such that $k(1+\log(p/k))$ is large compare to *n* (Verzelen, 2012). For our panel, with n=279 genotypes and p=60k SNPs, this threshold is reached around k=30.

To circumvent the fact that the methods account for LD differently, we compared all methods in terms of QTLs, defined here as intervals around significant SNPs, instead of significant SNPs directly. But even here the fact that we used the genome-wide distribution of LD to define the extent of QTLs ignores local variations of LD along the genome. Adding haplotype-based methods to the comparison could provide complementary information (Lorenz *et al.*, 2010), but is beyond the scope of this work as it requires first to infer local haplotypes, a difficult endeavor in itself, especially for highly heterozygous individuals, and then to account for haplotype uncertainty when testing the null hypothesis of no association between the haplotype and the response.

We compared our QTLs only with those from the literature which passed a genome-wide significance threshold. When we deemed one of our QTL to be new, it may nevertheless have been found in a bi-parental cross at the chromosome-wide significance threshold. Furthermore, such a comparison could be achieved only for a very small subset of traits. Part of the reason why may be publication bias (Rothstein *et al.*, 2005): many traits were analyzed with the interval-mapping method but only those with at least one QTL were mentioned in publications. In addition, we were faced with the notorious difficulty to assess if the same trait acronym used in different articles indeed corresponded to the same biological trait. A wider usage of a trait ontology, such as the *Vitis* ontology, to harness QTL results across studies seems the way forward (Krajewski *et al.*, 2015).

When comparing our QTLs with genomic annotations, we did find hundreds of hits. Beyond those already known (e.g., on chr2 around the MYB genes for anthocyanin-related response variables, Matus *et al.*, 2008), we hope such a database will help in refining existing annotations and suggesting new ones, as aimed in the INTEGRAPE initiative (http://www.integrape.eu). Ultimately, this should help prioritizing candidate genes for follow-up studies.

Genomic prediction, and the wider goal of understanding genetic architectures

The accuracy of genomic prediction, when assessed by cross-validation within the panel, reached promising levels: the median Pearson correlation around 0.4 corresponds to a moderately linear relationship between predicted and empirical genotypic BLUPs. This is notably the case for traits displaying a high broad-sense heritability, but not always. Genomic prediction can hence be useful for traits hard to measure accurately. In parallel, the coefficient of determination remains substantially lower (around 0.17), indicating that the variation of predicted genotypic BLUPs only explains a small proportion of the variance in empirical genotypic BLUPs. Nevertheless, in

25

694 selection, one mostly cares in accurately predicting the ranks of candidate genotypes, and the
695 median Spearman correlation around 0.4 is relevant in that case.
696 Cross-validation results are interesting *per se* as they provide an upper threshold on
697 prediction accuracy. Yet, the ultimate goal lies in training a model on a panel to predict genotypic
698 BLUPs in a segregation population. When genomic prediction for mean berry weight was
699 performed on a progeny, *i.e.*, on genotypes not part of the panel, the accuracy was lower than the
700 results obtained by within-panel cross-validation, yet they displayed the same trend in terms of
701 methods. The fact that the ridge regression model (rrBLUP) performed better than the sparse
702 regression model (varbvs) may be due to both the infinitesimal architecture of the trait as well as the
703 lack of segregating QTLs for this trait in the progeny. This promising result now needs to be
704 confirmed with other traits and other, more complex progenies in a future work, in the same spirit as
705 what was done on other perennial fruit crops (Muranty *et al.*, 2015; Minamikawa *et al.*, 2017).
706 Indeed, genomic prediction in perennial crops is known to be promising (e.g., Grattapaglia and
707 Resende, 2011), "as long as models are used at the relevant selection age and within the breeding
708 zone in which they were estimated" (Resende *et al.*, 2012).
709 Furthermore, this diverse panel of 279 *Vitis vinifera* L. could represent the main building
710 block of an international consortium in construction gathering geneticists, physiologists,
711 biochemists, modelers and breeders working on grapevine as discussed during the Grapevine
712 Breeding and Genetics in 2018 in Bordeaux, France. For instance, to study genotype-environment
713 interactions in the vineyard, several research groups pledged to plant the panel in two randomized
714 blocks at their site. This will notably allow to study the genetic basis of various phenological traits
715 on the same plant material in contrasted sites. Other research groups are invited to contact us for
716 more details. In parallel, the panel will be studied for traits related to drought in more controlled
717 environments, extending what was done on a bi-parental cross (Coupel-Ledru *et al.*, 2014, 2016).
718 Beyond the results on individual QTLs from GWAS and on overall accuracy from
719 genomic prediction, our study also aimed at providing basic insights into the genetic architecture of
720 various traits of interest for grapevine. In this goal, we initially used a Bayesian sparse linear mixed
721 model, BSLMM (Zhou *et al.*, 2013), as it includes both the Bayesian variable selection regression
722 and the ridge regression as special cases. However, likely due to the small size of our panel compare
723 to the data sets analyzed in the original article, the parameter uncertainty was too high to be
724 meaningfully interpreted (Flutre *et al.*, 2018). Nevertheless, we took advantage of the large number
725 of diverse traits, all analyzed in the same way, to shed some light on the interplay between the
726 accuracy with which phenotypic measurements translates into genotypic values, the number of
727 QTLs that can be reliably detected, and the differentiated prediction accuracy depending on

assumptions about the underlying genetic architectures. In our analyzes, we focused on the part of genetic architectures restricted to the additive genetic variance because including dominance genetic variance led to convergence issues. But more generally, strong arguments exist in favor of focusing only on the additive part (Hill *et al.*, 2008). In this context, the key difference between genetic architectures lies between the infinitesimal and the sparse architectures, and has been amply studied, *e.g.*, Daetwyler *et al.* (2010) and Wimmer *et al.* (2013). However, these articles focused on simulations or only analyzed annual crops for a small number of traits.

Our contribution on this topic confirmed the importance of heritability to detect QTLs and predict accurately. Indeed, detecting very few (or even no) QTLs for a given trait for which there is substantial genetic variance, could be interpreted as an absence of QTL with a strong-enough effect to be significant, hence as an indication of the genetic architecture being infinitesimal. In contrast, detecting several QTLs could suggest a sparse architecture. Nevertheless, as always with real data compared to simulations, it can also mean that the empirical BLUPs of the genotypic values are too noisy versions of the true genotypic values, hence no reliable QTL can be significantly detected, whatever the genetic architecture. Coloring points as in Figure 5 with respect to broad-sense heritability highlighted the importance of this metric when interpreting the relationship between the other two (difference in prediction accuracy and number of reliable QTLs). As a practical consequence, for response variables with a low broad-sense heritability, it seems more judicious to use a model assuming an infinitesimal architecture.

In the case of traits with low heritability, our results were in agreement with Wimmer *et al.* (2013) to recommend using the ridge regression BLUP, even though the genetic architecture underlying such traits is not infinitesimal. But most importantly, in contrast to Wimmer *et al.* (2013), we found many traits for which a variable selection method did predict better than the ridge regression BLUP, even though our sample size remained very low compared to studies on farm animals and humans. This may be due to the fact that we studied a perennial crop in which linkage disequilibrium falls very quickly compared to the long-range LD in annual crops studied by Wimmer *et al.* (2013). In the end, for breeding purposes, it may be sufficient to use a robust method such as the ridge regression whatever the trait. However, in basic research, we recommend to compare at least two methods, one assuming the infinitesimal model and another assuming a sparse architecture, and to put the results in perspective using estimates of heritability.

27

Acknowledgments

For funding: GrapeReSeq (ANR, 2009-2011), DLVitis (ANR, 2010-2012), Innovine (KBBE, 2014-2015), “Créer les cépages de demain avec les outils d’aujourd’hui” (CASDAR, 2011-2013), FruitSelGen (INRA méta-programme Selgen, 2015-2016). For the phenotyping of organic acids: Valérie Miralles and Jean-François Ballester from AGAP-PPB. For the sequencing: Pierre Mournet from AGAP-GPTR, GenoToul. For the computing: Bertrand Pitollat and the South Green platform.

Author contribution

PT, AD, JMB and LLC initiated the project. AD and JPP conceived the experimental design in the field. GB and YB installed and managed the field trial under the supervision of JPP and LLC. GB, YB, JPP, AD, LLC, RB, TL, JMB, VL, PT collected phenotypic data on clusters and berries in 2010-2012. CR and LLC collected organic acid data from berries in 2011-2012. LLC, VC and JPP conceived the experimental design in 2014-2015. AF, GB, YB and LLC collected phenotyping data in 2014-2015 and extracted DNA samples for the first GBS phase. IB tested the presence of viruses. MR, GB, YB and LLC prepared samples before polyphenols, β-damascenone and pDMS extraction. VB collected β-damascenone and pDMS data. LLC and TF conceived the experimental design for the GBS. AL extracted DNA samples for the second GBS phase and made the libraries. TF wrote all the code and performed the analyzes. TF, AD and CR interpreted the results. TF drafted the manuscript. All authors contributed critical revision of the work and approved the manuscript.

References

Adam-Blondon A-F, Martínez-Zapater JM, Kole C (Eds.). 2011. *Genetics, genomics and breeding of grapes*.

Adam-Blondon A, Roux C, Claux D, Butterlin G, Merdinoglu D, This P. **2004**. Mapping 245 SSR markers on the Vitis vinifera genome: a tool for grape genetics. *Theoretical and Applied Genetics* **109**: 1017–1027.

Albrechtsen A, Nielsen FC, Nielsen R. **2010**. Ascertainment Biases in SNP Chips Affect Measures of Population Divergence. *Molecular Biology and Evolution* **27**: 2534–2547.

Arlot S, Lerasle M. **2016**. Choice of V for V-fold Cross-validation in Least-squares Density Estimation. *J. Mach. Learn. Res.* **17**: 7256–7305.

Astle W, Balding D. **2009**. Population structure and cryptic relatedness in genetic association studies. *Statistical Science* **24**: 451–471.

Baey C. **2014**. Modélisation de la variabilité inter-individuelle dans les modèles de croissance de plantes et sélection de modèles pour la prévision.

Barba P, Cadle-Davidson L, Harriman J, Glaubitz J, Brooks S, Hyma K, Reisch B. **2014**. Grapevine powdery mildew resistance and susceptibility loci identified on a high-resolution SNP map. *Theoretical and Applied Genetics* **127**: 73–84.

Bates D, Mächler M, Bolker B, Walker S. **2015**. Fitting linear mixed-effects models using lme4. *Journal of Statistical Software* **67**.

Battilana J, Costantini L, Emanuelli F, Sevini F, Segala C, Moser S, Velasco R, Versini G, Grando MS. **2009**. The 1-deoxy-d-xylulose 5-phosphate synthase gene co-localizes with a major QTL affecting monoterpene content in grapevine. *Theoretical and Applied Genetics* **118**: 653–669.

Bigard A, Berhe DT, Maoddi E, Sire Y, Boursiquot J-M, Ojeda H, Péros J-P, Doligez A, Romieu C, Torregrosa L. **2018**. Vitis vinifera L. Fruit Diversity to Breed Varieties Anticipating Climate Changes. *Frontiers in Plant Science* **9**.

Bonnafous F, Fievet G, Blanchet N, Boniface M-C, Carrère S, Gouzy J, Legrand L, Marage G, Bret-Mestries E, Munos S, *et al.* 2018. Comparison of GWAS models to identify non-additive genetic control of flowering time in sunflower hybrids. *Theoretical and Applied Genetics* **131**: 319–332.

Box G, Cox D. **1964**. An analysis of transformations. *Journal of the Royal Statistical Society. Series B (Methodological)* **26**: 211–252.

Browning BL, Browning SR. **2009**. A unified approach to genotype imputation and haplotype-phase inference for large data sets of trios and unrelated individuals. *The American Journal of Human Genetics* **84**: 210–223.

Burbidge JB, Magee L, Robb AL. **1988**. Alternative Transformations to Handle Extreme Values of the Dependent Variable. *Journal of the American Statistical Association* **83**: 123.

29

de los Campos G, Hickey JM, Pong-Wong R, Daetwyler HD, Calus MPL. **2013**. Whole-Genome Regression and Prediction Methods Applied to Plant and Animal Breeding. *Genetics* **193**: 327–345.

Canaguier A, Grimplet J, Di Gaspero G, Scalabrin S, Duchêne E, Choisne N, Mohellibi N, Guichard C, Rombauts S, Le Clainche I, *et al.* 2017. A new version of the grapevine reference genome assembly (12X.v2) and of its annotation (VCost.v3). *Genomics Data* **14**: 56–62.

Carbonetto P, Stephens M. **2012**. Scalable variational inference for Bayesian variable selection in regression, and its accuracy in genetic association studies. *Bayesian Analysis* **7**: 73–108.

Cardon LR, Bell JI. **2001**. Association study designs for complex diseases. *Nature Reviews Genetics* **2**: 91–99.

Carpenter J, Bithell J. **2000**. Bootstrap confidence intervals: when, which, what? A practical guide for medical statisticians. *Statistics in Medicine* **19**: 1141–1164.

Chen J, Chen Z. **2008**. Extended Bayesian information criteria for model selection with large model spaces. *Biometrika* **95**: 759–771.

Clark MF, Adams AN. **1977**. Characteristics of the Microplate Method of Enzyme-Linked Immunosorbent Assay for the Detection of Plant Viruses. *Journal of General Virology* **34**: 475–483.

Coupel-Ledru A, Lebon É, Christophe A, Doligez A, Cabrera-Bosquet L, Péchier P, Hamard P, This P, Simonneau T. **2014**. Genetic variation in a grapevine progeny (Vitis vinifera L. cvs Grenache×Syrah) reveals inconsistencies between maintenance of daytime leaf water potential and response of transpiration rate under drought. *Journal of Experimental Botany*: eru228.

Coupel-Ledru A, Lebon E, Christophe A, Gallo A, Gago P, Pantin F, Doligez A, Simonneau T. **2016**. Reduced nighttime transpiration is a relevant breeding target for high water-use efficiency in grapevine. *Proceedings of the National Academy of Sciences* **113**: 8963–8968.

Daetwyler H, Pong-Wong R, Villanueva B, Woolliams J. **2010**. The Impact of Genetic Architecture on Genome-Wide Evaluation Methods. *Genetics* **185**: 1021–1031.

Di Gaspero G, Cipriani G, Adam-Blondon A-F, Testolin R. **2007**. Linkage maps of grapevine displaying the chromosomal locations of 420 microsatellite markers and 82 markers for R-gene candidates. *Theoretical and Applied Genetics* **114**: 1249–1263.

Doligez A, Bertrand Y, Farnos M, Grolier M, Romieu C, Esnault F, Dias S, Berger G, François P, Pons T, *et al.* 2013. New stable QTLs for berry weight do not colocalize with QTLs for seed traits in cultivated grapevine (Vitis vinifera L.). *BMC Plant Biology* **13**: 1–16.

Duchêne E. **2020**. Vitis INRAE ontology. https://urgi.versailles.inra.fr/ephesis/ephesis/ontologyportal.do

Duchêne E, Butterlin G, Claudel P, Dumas V, Jaegli N, Merdinoglu D. **2009**. A grapevine (Vitis vinifera L.) deoxy-d-xylulose synthase gene colocates with a major quantitative trait loci for terpenol content. *Theoretical and Applied Genetics* **118**: 541–552.

30

Elshire R, Glaubitz J, Sun Q, Poland J, Kawamoto K, Buckler E, Mitchell S. **2011**. A Robust, Simple Genotyping-by-Sequencing (GBS) Approach for High Diversity Species. *PLoS One* **6**: e19379.

Endelman J. **2011**. Ridge regression and other kernels for genomic selection with R package rrBLUP. *The Plant Genome Journal* **4**: 250.

Falush D, Stephens M, Pritchard J. **2003**. Inference of Population Structure Using Multilocus Genotype Data: Linked Loci and Correlated Allele Frequencies. *Genetics* **164**: 1567–1587.

Fiorani F, Schurr U. **2013**. Future Scenarios for Plant Phenotyping. *Annual Review of Plant Biology* **64**: 267–291.

Flutre T. **2018**. Genome-wide association study of a diverse grapevine panel to uncover the genetic architecture of numerous traits of interest.

Fournier-Level A, Le Cunff L, Gomez C, Doligez A, Ageorges A, Roux C, Bertrand Y, Souquet J-M, Cheynier V, This P. **2009**. Quantitative genetic bases of anthocyanin variation in grape (Vitis vinifera L. ssp. sativa) berry: a quantitative trait locus to quantitative trait nucleotide integrated study. *Genetics* **183**: 1127–1139.

Gardner M, Altman D. **1986**. Confidence intervals rather than P values: estimation rather than hypothesis testing. *Br Med J (Clin Res Ed)* **292**: 746–750.

Gaudillère J-P, Van Leeuwen C, Trégoat O. **2001**. The assessment of vine water uptake conditions by $^{13}c/^{12}c$ discrimination in grape sugar. *OENO One* **35**: 195.

Génard M, Bertin N, Gautier H, Lescourret F, Quilot B. **2010**. Virtual profiling: a new way to analyse phenotypes: Virtual profiling to analyse phenotypes. *The Plant Journal* **62**: 344–355.

Gianola D, Cecchinato A, Naya H, Schön C-C. **2018**. Prediction of Complex Traits: Robust Alternatives to Best Linear Unbiased Prediction. *Frontiers in Genetics* **9**.

Goodman SN, Fanelli D, Ioannidis JPA. **2016**. What does research reproducibility mean? *Science Translational Medicine* **8**: 341ps12-341ps12.

Grattapaglia D, Resende M. **2011**. Genomic selection in forest tree breeding. *Tree Genetics & Genomes* **7**: 241–255.

Guo D-L, Zhao H-L, Li Q, Zhang G-H, Jiang J-F, Liu C-H, Yu Y-H. **2019**. Genome-wide association study of berry-related traits in grape [Vitis vinifera L.] based on genotyping-by-sequencing markers. *Horticulture Research* **6**.

Gurka MJ, Edwards LJ, Muller KE, Kupper LL. **2006**. Extending the Box-Cox transformation to the linear mixed model. *Journal of the Royal Statistical Society: Series A (Statistics in Society)* **169**: 273–288.

Habier D, Fernando R, Dekkers J. **2007**. The impact of genetic relationship information on genome-assisted breeding values. *Genetics* **177**: 2389–2397.

Hill W, Goddard M, Visscher P. **2008**. Data and Theory Point to Mainly Additive Genetic Variance for Complex Traits. *PLoS Genetics* **4**: e1000008.

31

Hoggart C, Whittaker J, De Iorio M, Balding D. **2008**. Simultaneous analysis of all SNPs in genome-wide and re-sequencing association studies. *PLoS Genetics* **4**: e1000130.

Jacquin L, Cao T-V, Ahmadi N. **2016**. A unified and comprehensible view of parametric and kernel methods for genomic prediction with application to rice. *Frontiers in Genetics* **7**.

Jung H, Winefield C, Bombarely A, Prentis P, Waterhouse P. **2019**. Tools and Strategies for Long-Read Sequencing and De Novo Assembly of Plant Genomes. *Trends in Plant Science* **24**: 700–724.

Kicherer A, Herzog K, Bendel N, Klück H-C, Backhaus A, Wieland M, Rose J, Klingbeil L, Läbe T, Hohl C, *et al.* 2017. Phenoliner: A New Field Phenotyping Platform for Grapevine Research. *Sensors* **17**: 1625.

Klein LL, Miller AJ, Ciotir C, Hyma K, Uribe-Convers S, Londo J. **2018**. High-throughput sequencing data clarify evolutionary relationships among North American *Vitis* species and improve identification in USDA *Vitis* germplasm collections. *American Journal of Botany* **105**: 215–226.

Kotseridis Y, Baumes RL, Skouroumounis GK. **1999**. Quantitative determination of free and hydrolytically liberated β-damascenone in red grapes and wines using a stable isotope dilution assay. *Journal of Chromatography A* **849**: 245–254.

Krajewski P, Chen D, Ćwiek H, van Dijk A, Fiorani F, Kersey P, Klukas C, Lange M, Markiewicz A, Nap J, *et al.* 2015. Towards recommendations for metadata and data handling in plant phenotyping. *Journal of Experimental Botany* **66**: 5417–5427.

Kruglyak L. **1999**. Prospects for whole-genome linkage disequilibrium mapping of common disease genes. *Nature Genetics* **22**: 139–144.

Kuhn M. **2018**. *caret: Classification and Regression Training*.

Kuznetsova A, Brockhoff PB, Christensen RHB. **2017**. **lmerTest** Package: Tests in Linear Mixed Effects Models. *Journal of Statistical Software* **82**.

Laucou V, Launay A, Bacilieri R, Lacombe T, Adam-Blondon A-F, Bérard A, Chauveau A, de Andrés MT, Hausmann L, Ibáñez J, *et al.* 2018. Extended diversity analysis of cultivated grapevine Vitis vinifera with 10K genome-wide SNPs (T-Y Chiang, Ed.). *PLOS ONE* **13**: e0192540.

Lorenz AJ, Hamblin MT, Jannink J-L. **2010**. Performance of Single Nucleotide Polymorphisms versus Haplotypes for Genome-Wide Association Analysis in Barley (I Baxter, Ed.). *PLoS ONE* **5**: e14079.

Marchini J, Howie B. **2010**. Genotype imputation for genome-wide association studies. *Nature Reviews Genetics* **11**: 499–511.

Mardia KV, Kent JT, Bibby JM. **1979**. *Multivariate analysis*. London ; New York: Academic Press.

Marguerit E, Boury C, Manicki A, Donnart M, Butterlin G, Némorin A, Wiedemann-Merdinoglu S, Merdinoglu D, Ollat N, Decroocq S. **2009**. Genetic dissection of sex determinism,

925 inflorescence morphology and downy mildew resistance in grapevine. *Theoretical and Applied*
926 *Genetics* **118**: 1261–1278.
927 **Marrano A, Birolo G, Prazzoli ML, Lorenzi S, Valle G, Grando MS**. **2017**. SNP-Discovery by
928 RAD-Sequencing in a Germplasm Collection of Wild and Cultivated Grapevines (V. vinifera L.) (S
929 Amancio, Ed.). *PLOS ONE* **12**: e0170655.
930 **Marroni F, Pinosio S, Morgante M**. **2014**. Structural variation and genome complexity: is
931 dispensable really dispensable? *Current Opinion in Plant Biology* **18**: 31–36.
932 **Matus J, Aquea F, Arce-Johnson P**. **2008**. Analysis of the grape MYB R2R3 subfamily reveals
933 expanded wine quality-related clades and conserved gene structure organization across Vitis and
934 Arabidopsis genomes. *BMC Plant Biology* **8**: 83.
935 **McShane BB, Gal D**. **2017**. Statistical Significance and the Dichotomization of Evidence. *Journal*
936 *of the American Statistical Association* **112**: 885–895.
937 **Mejía N, Soto B, Guerrero M, Casanueva X, Houel C, de los Ángeles Miccono M, Ramos R,**
938 **Le Cunff L, Boursiquot J-M, Hinrichsen P, *et al.* 2011**. Molecular, genetic and transcriptional
939 evidence for a role of VvAGL11 in stenospermocarpic seedlessness in grapevine. *BMC Plant*
940 *Biology* **11**: 57.
941 **Meuwissen T, Hayes B, Goddard M**. **2001**. Prediction of total genetic value using genome-wide
942 dense marker maps. *Genetics* **157**: 1819–1829.
943 **Migicovsky Z, Sawler J, Gardner KM, Aradhya MK, Prins BH, Schwaninger HR,**
944 **Bustamante CD, Buckler ES, Zhong G-Y, Brown PJ, *et al.* 2017**. Patterns of genomic and
945 phenomic diversity in wine and table grapes. *Horticulture Research* **4**: 17035.
946 **Minamikawa MF, Nonaka K, Kaminuma E, Kajiya-Kanegae H, Onogi A, Goto S, Yoshioka T,**
947 **Imai A, Hamada H, Hayashi T, *et al.* 2017**. Genome-wide association study and genomic
948 prediction in citrus: Potential of genomics-assisted breeding for fruit quality traits. *Scientific*
949 *Reports* **7**.
950 **Möhring J, Piepho H-P**. **2009**. Comparison of Weighting in Two-Stage Analysis of Plant Breeding
951 Trials. *Crop Science* **49**: 1977.
952 **Molnar M, Ilie L**. **2015**. Correcting Illumina data. *Briefings in Bioinformatics* **16**: 588–599.
953 **Muranty H, Troggio M, Sadok IB, Rifaï MA, Auwerkerken A, Banchi E, Velasco R, Stevanato**
954 **P, van de Weg WE, Di Guardo M, *et al.* 2015**. Accuracy and responses of genomic selection on
955 key traits in apple breeding. *Horticulture Research* **2**.
956 **Myles S, Boyko A, Owens C, Brown P, Grassi F, Aradhya M, Prins B, Reynolds A, Chia J-M,**
957 **Ware D, *et al.* 2011**. Genetic structure and domestication history of the grape. *Proceedings of the*
958 *National Academy of Sciences* **108**: 3530–3535.
959 **Myles S, Peiffer J, Brown PJ, Ersoz ES, Zhang Z, Costich DE, Buckler ES**. **2009**. Association
960 Mapping: Critical Considerations Shift from Genotyping to Experimental Design. *The Plant Cell*
961 **21**: 2194–2202.

Nanson A. **1970**. L'héritabilité et le gain d'origine génétique dans quelques types d'expériences. *Silvae Genetica* **19**: 113–121.

Negi SS, Olmo HP. **1966**. Sex conversion in a male Vitis vinifera L. by a kinin. *Science* **152**: 1624–1624.

Nicolas S, 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 Biology* **16**.

Oakey H, Verbyla A, Pitchford W, Cullis B, Kuchel H. **2006**. Joint modeling of additive and non-additive genetic line effects in single field trials. *Theoretical and Applied Genetics* **113**: 809–819.

Pallas B, Loi C, Christophe A, Cournède P-H, Lecoeur J. **2009**. A Stochastic Growth Model of Grapevine with Full Interaction Between Environment, Trophic Competition and Plant Development. In: IEEE, 95–102.

Pérez P, Gustavo de los Campos. **2014**. Genome-Wide Regression and Prediction with the BGLR Statistical Package. *Genetics* **198**: 483–495.

Picq S, Santoni S, Lacombe T, Latreille M, Weber A, Ardisson M, Ivorra S, Maghradze D, Arroyo-Garcia R, Chatelet P, *et al.* 2014. A small XY chromosomal region explains sex determination in wild dioecious V. vinifera and the reversal to hermaphroditism in domesticated grapevines. *BMC Plant Biology* **14**.

Pinasseau L, Vallverdú-Queralt A, Verbaere A, Roques M, Meudec E, Le Cunff L, Péros J-P, Ageorges A, Sommerer N, Boulet J-C, *et al.* 2017a. Cultivar Diversity of Grape Skin Polyphenol Composition and Changes in Response to Drought Investigated by LC-MS Based Metabolomics. *Frontiers in Plant Science* **8**.

Pinasseau L, Verbaere A, Roques M, Meudec E, Vallverdu-Queralt A, Ollier L, Marlin T, Guiraud J-L, Berger G, Bertrand Y, *et al.* 2017b. Innovine WP3: 105 phenolic compound quantification of 2014 and 2015 mature grape berries from a core-collection of 279 irrigated and non-irrigated Vitis vinifera cultivars.

Piñeiro G, Perelman S, Guerschman JP, Paruelo JM. **2008**. How to evaluate models: Observed vs. predicted or predicted vs. observed? *Ecological Modelling* **216**: 316–322.

Pritchard J, Stephens M, Donnelly P. **2000**. Inference of population structure using multilocus genotype data. *Genetics* **155**: 945–959.

Resende MFR, Muñoz P, Acosta JJ, Peter GF, Davis JM, Grattapaglia D, Resende MDV, Kirst M. **2012**. Accelerating the domestication of trees using genomic selection: accuracy of prediction models across ages and environments. *New Phytologist* **193**: 617–624.

Rex F, Fechter I, Hausmann L, Töpfer R. **2014**. QTL mapping of black rot (Guignardia bidwellii) resistance in the grapevine rootstock 'Börner' (V. riparia Gm183 × V. cinerea Arnold). *Theoretical and Applied Genetics* **127**: 1667–1677.

Reynolds D, Baret F, Welcker C, Bostrom A, Ball J, Cellini F, Lorence A, Chawade A, Khafif M, Noshita K, *et al.* 2019. What is cost-efficient phenotyping? Optimizing costs for different scenarios. *Plant Science* **282**: 14–22.

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 Biology* **16**.

Rockman MV. 2012. The QTN program and the alleles that matter for evolution: all that's gold does not glitter. *Evolution* **66**: 1–17.

Rothstein HR, Sutton AJ, Borenstein M (Eds.). 2005. *Publication Bias in Meta-Analysis: Prevention, Assessment and Adjustments*. Chichester, UK: John Wiley & Sons, Ltd.

Schweiger R, Kaufman S, Laaksonen R, Kleber ME, März W, Eskin E, Rosset S, Halperin E. 2016. Fast and accurate construction of confidence intervals for heritability. *The American Journal of Human Genetics* **98**: 1181–1192.

Segura V, Vilhjalmsson B, Platt A, Korte A, Seren U, Long Q, Nordborg M. 2012. An efficient multi-locus mixed-model approach for genome-wide association studies in structured populations. *Nature Genetics* **44**: 825–830.

Segurel MA, Razungles AJ, Riou C, Trigueiro MGL, Baumes RL. 2005. Ability of Possible DMS Precursors To Release DMS during Wine Aging and in the Conditions of Heat-Alkaline Treatment. *Journal of Agricultural and Food Chemistry* **53**: 2637–2645.

Shahood R. 2017. La baie au sein d'une vendange asynchrone : Un nouveau paradigme vers l'interprétation quantitative des flux de sucres et acides en tant qu'osmoticums et substrat respiratoires majeurs lors du développement bimodal du raisin.

Sievanen R, Godin C, DeJong TM, Nikinmaa E. 2014. Functional-structural plant models: a growing paradigm for plant studies. *Annals of Botany* **114**: 599–603.

Swarts K, Li H, Alberto Romero Navarro, An D, Romay M, Hearne S, Acharya C, Glaubitz J, Mitchell S, Elshire R, *et al.* 2014. Novel Methods to Optimize Genotypic Imputation for Low-Coverage, Next-Generation Sequence Data in Crop Plants. *The Plant Genome* **7**.

VanRaden P. 2008. Efficient methods to compute genomic predictions. *Journal of Dairy Science* **91**: 4414–4423.

Verzelen N. 2012. Minimax risks for sparse regressions: Ultra-high dimensional phenomenons. *Electronic Journal of Statistics* **6**: 38–90.

Vezzulli S, Zulini L, Stefanini M. 2019. Genetics-assisted breeding for downy/powdery mildew and phylloxera resistance at fem (J-M Aurand, Ed.). *BIO Web of Conferences* **12**: 01020.

Vitezica Z, Varona L, Legarra A. 2013. On the additive and dominant variance and covariance of individuals within the genomic selection scope. *Genetics* **195**: 1223–1230.

Wang X, Yang Z, Xu C. 2015. A comparison of genomic selection methods for breeding value prediction. *Science Bulletin* **60**: 925–935.

Wimmer V, Lehermeier C, Albrecht T, Auinger H-J, Wang Y, Schon C-C. **2013**. Genome-Wide Prediction of Traits with Different Genetic Architecture Through Efficient Variable Selection. *Genetics* **195**: 573–587.

Wolkovich EM, García de Cortázar-Atauri I, Morales-Castilla I, Nicholas KA, Lacombe T. **2018**. From Pinot to Xinomavro in the world's future wine-growing regions. *Nature Climate Change* **8**: 29–37.

Xu S. **2003a**. Estimating polygenic effects using markers of the entire genome. *Genetics* **163**: 789–801.

Xu S. **2003b**. Theoretical basis of the Beavis effect. *Genetics* **165**: 2259–2268.

Yang J, Benyamin B, McEvoy B, Gordon S, Henders A, Nyholt D, Madden P, Heath A, Martin N, Montgomery G, *et al.* 2010. Common SNPs explain a large proportion of the heritability for human height. *Nature genetics* **42**: 565–569.

Yang X, Guo Y, Zhu J, Niu Z, Shi G, Liu Z, Li K, Guo X. **2017**. Genetic Diversity and Association Study of Aromatics in Grapevine. *Journal of the American Society for Horticultural Science* **142**: 225–231.

Zarouri B. **2016**. Association study of phenology, yield and quality related traits in table grapes using SSR and SNP markers.

Zhang H, Fan X, Zhang Y, Jiang J, Liu C. **2017**. Identification of favorable SNP alleles and candidate genes for seedlessness in Vitis vinifera L. using genome-wide association mapping. *Euphytica* **213**.

Zhang Y-M, Jia Z, Dunwell JM. **2019**. Editorial: The Applications of New Multi-Locus GWAS Methodologies in the Genetic Dissection of Complex Traits. *Frontiers in Plant Science* **10**: 100.

Zhou X, Carbonetto P, Stephens M. **2013**. Polygenic modeling with Bayesian sparse linear mixed models. *PLoS Genetics* **9**: e1003264.

Zhou X, Stephens M. **2012**. Genome-wide efficient mixed-model analysis for association studies. *Nature Genetics* **44**: 821–824.

36

Supporting information

Fig. S1 Layout of the field trial

Fig. S2 Distribution of raw phenotypic data per response variable, block and year

Fig. S3 Variogram of controls' residuals per response variable and year

Fig. S4 Principal component analysis with 8840 microarray-only SNPs and assignments from Nicolas *et al.* (2016)

Fig. S5 Distribution of minor allele frequencies of SNPs per SNP set

Fig. S6 SNP density per chromosome per SNP set

Fig. S7 Additive and dominance genetic relationships per SNP set

Fig. S8 Physical distance between SNP pairs plotted against their linkage disequilibrium value per SNP set

Fig. S9 Numbers of QTLs found by one, two, or the three statistical methods

Fig. S10 Numbers of response variables with at least one QTL found by one, two, or the three statistical methods

Fig. S11 Genomic distribution of QTLs per statistical method

Fig. S12 Genomic distribution of reliable QTLs, found with at least two statistical methods

Fig. S13 Distributions of genotypic values per genotypic class for one significant SNP per QTL, for all reliable QTLs

Fig. S14 Spearman correlation coefficient (from cross-validation) per response variable plotted against estimated broad-sense heritabilities, per method and SNP set

Fig. S15 Pearson correlation coefficients between the empirical BLUPs of genotypic values for all pairs of response variables

Table S1 Barcodes used to multiplex samples before sequencing

Table S2 List of traits along with the fixed and random effects of the final linear mixed model selected for each response variable, as well as the estimates and confidence intervals of broad-sense heritability and coefficient of genetic variation

Table S3 List of cultivars with genetic assignment modified between SSRs and microarray SNPs

Table S4 Estimates and confidence intervals of narrow-sense heritability (as the proportion of variance in genotypic BLUPs explained by SNPs, PVE_{SNPs}) per response variable, by maximum

37

likelihood (ML) and restricted maximum likelihood (ReML), using additive-only and additive+dominance variance components

Table S5 List of significant SNPs per SNP set and statistical method

Table S6 List of significant QTLs per SNP set and statistical method

Table S7 List of significant QTLs merged per response variable over statistical methods and SNP sets, along with their reliability

Table S8 Classification of the traits according to the *Vitis* ontology

Table S9 List of reliable QTLs merged over response variables

Table S10 List of positional candidate gene models for the reliable QTLs merged over response variables

Table S11 Annotations of positional candidate genes for the reliable QTLs merged over response variables

Table S12 Average performance metrics from the repeated K-fold cross-validation performed on all response variables per SNP set and statistical method

Table S13 Broad-sense heritability, number of reliable QTLs and difference between averaged Spearman correlations of varbvs and rrBLUP from crossvalidation per response variable

Table S14 Matrix of Pearson correlation coefficients between the empirical BLUPs of genotypic values for all pairs of response variables

38